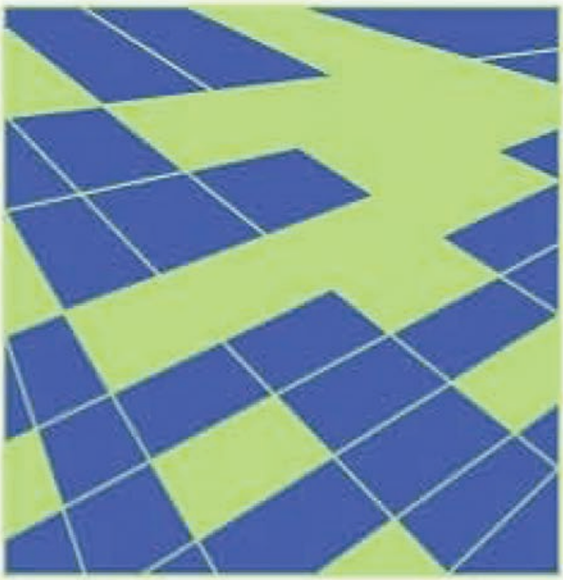
Cambridge Studies in Probability, Induction, and Decision Theory

KNOWLEDGE AND INQUIRY

Essays on the Pragmatism of Isaac Levi



Edited by

ERIK J. OLSSON

Knowledge and Inquiry

Isaac Levi, John Dewey Professor of Philosophy emeritus at Columbia University, has explored the principles of American pragmatism in greater depth and more consistency than others before him. The result is a sophisticated and powerful philosophical system whose key elements stand in stark opposition not only to current mainstream epistemology, but also to the positions of other contemporary authors writing in the same pragmatist tradition. The essays in this timely volume, written by some of philosophy's finest scholars, contribute substantially to the understanding and appraisal of Levi's work. Included in this volume are Levi's extensive and provocative replies to his critics, which offer unique access to his current thinking on a wide range of topics. The introduction provides a concise, systematic presentation of the cornerstone of Levi's pragmatism. Suitable for students and scholars who are interested in American pragmatism in general and Isaac Levi's work in particular, this book is an ideal companion to Levi's own writings.

Erik J. Olsson is senior lecturer in the department of philosophy at Lund University, Sweden. He has published extensively on epistemology, philosophy of science, and logic, and he is the author of *Against Coherence*.

Cambridge Studies in Probability, Induction, and Decision Theory

General editor: Brian Skyrms

Ellery Eells, *Probabilistic Causality*
Richard Jeffrey, *Probability and the Art of Judgment*
Robert C. Koons, *Paradoxes of Belief and Strategic Rationality*
Cristina Bicchieri and Maria Luisa Dalla Chiara (eds.), *Knowledge, Belief, and Strategic Interactions*
Patrick Maher, *Betting on Theories*
Cristina Bicchieri, *Rationality and Coordination*
J. Howard Sobel, *Taking Chances*
Jan von Plato, *Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective*
Ellery Eells and Brian Skyrms (eds.), *Probability and Conditionals*
Cristina Bicchieri, Richard Jeffrey, and Brian Skyrms (eds.), *The Dynamics of Norms*
Patrick Suppes and Mario Zanotti, *Foundations of Probability with Applications*
Paul Weirich, *Equilibrium and Rationality*
Daniel Hausman, *Causal Asymmetries*
William A. Dembski, *The Design Inference*
James M. Joyce, *The Foundations of Causal Decision Theory*
Yair Guttmann, *The Concept of Probability in Statistical Physics*
Joseph B. Kadane, Mark B. Schervish, and Teddy Seidenfeld (eds.), *Rethinking the Foundations of Statistics*
Phil Dowe, *Physical Causation*

Knowledge and Inquiry

Essays on the Pragmatism of Isaac Levi

Edited by

ERIK J. OLSSON

Lund University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
40 West 20th Street, New York, NY 10011-4211, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521845564

First published 2006

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Knowledge and inquiry : essays on the pragmatism of Isaac Levi /
edited by Erik J. Olsson.
p. cm. – (Cambridge studies in probability, induction, and decision theory)
Includes bibliographical references and index.
ISBN-0-521-84556-4
1. Pragmatism – History – 20th century. 2. Levi, Isaac, 1930–
I. Olsson, Erik J. II. Series.
B944.P72K56 2006
191–dc22 2005010532

ISBN-13 978-0-521-84556-4 hardback
ISBN-10 0-521-84556-4 hardback

I dedicate this book to
my mother, Mary Olsson.

Contents

Contributors

Horacio Arló Costa
Department of Philosophy
Carnegie Mellon University
Pittsburgh

Akeel Bilgrami
Department of Philosophy
Columbia University
New York

Otávio Bueno
Department of Philosophy
University of South Carolina
Columbia

André Fuhrmann
Department of Philosophy
University of São Paulo

Bengt Hansson
Department of Philosophy
Lund University

Sven Ove Hansson
Philosophy Unit
Royal Institute of Technology
Stockholm

Wolfram Hinzen
Department of Philosophy
University of Amsterdam

Mark Kaplan
Department of Philosophy
Indiana University
Bloomington

Philip Kitcher
Department of Philosophy
Columbia University
New York

Henry E. Kyburg, Jr.
University of Rochester and
The Institute of Human and Machine Cognition
University of West Florida
Pensacola

Isaac Levi
Department of Philosophy
Columbia University
New York

D. H. Mellor
Faculty of Philosophy
Cambridge University

Cheryl Misak
Department of Philosophy
University of Toronto

Erik J. Olsson
Department of Philosophy
Lund University

Maurice Pagnucco
School of Computer Science
and Engineering
The University of New South Wales
Sydney

Johannes Persson
Department of Philosophy
Lund University

Włodek Rabinowicz
Department of Philosophy
Lund University

Hans Rott
Institute for Philosophy
Regensburg University

Nils-Eric Sahlin
Department of Philosophy
Lund University

Wolfgang Spohn
Department of Philosophy
Constance University

Preface

Isaac Levi has retired from his academic position as prestigious John Dewey Professor of Philosophy at Columbia University, and yet there is no sign that he has retired from philosophy. On the contrary, he just published a new book entitled *Mild Contraction* with Oxford University Press, and he continues to write papers and participate in conferences and workshops. This collection is dedicated to Isaac on the occasion of his retirement, in celebration of his exceptional contribution to philosophy in general and to the great tradition of American pragmatism in particular. It is far more than a Festschrift in the usual sense; all the papers included are substantial contributions to the understanding and appraisal of Isaac's pragmatist philosophy by leading experts on his work. They were all written specifically for this volume and appear here for the first time. I am greatly indebted to all authors for their dedication and commitment to this project, and especially to Isaac for his extensive replies to all the papers.

My own intellectual debt to Isaac is great. My first internationally published paper, written jointly with Sven Ove Hansson, who was my thesis adviser back then, addressed Isaac's theory of belief contraction. In this connection, reading his book *The Fixation of Belief and Its Undoing* was especially rewarding, although some of his more intricate arguments were beyond my grasp at the time. Only gradually, on reading and rereading his books and articles, have I come to appreciate fully the extraordinary coherence and stringency of his thinking. I first encountered Isaac in Uppsala, Sweden, where he gave a lecture in the mid-1990s. Since then we have met and discussed philosophy on numerous occasions in the United States and in Europe, often attending the same conferences. He has always been extremely generous with his time and philosophical insight. As for personal qualities, I admire him above all for his intellectual honesty and integrity.

While the chapters in this book were being written, *Mild Contraction* had not yet appeared in print. I would like to thank Peter Momtchiloff and Rupert Cousens at Oxford University Press for making a preliminary electronic version available to the authors. The chapters in this volume are very much up to date, as they take Isaac's most recent work into account. It can be read as a companion volume to *Mild Contraction* or indeed to any other of Isaac's books. While some of the chapters contribute to technical aspects of Isaac's work, the overwhelming majority address more general philosophical issues, such as the nature of pragmatism and the role of truth in inquiry. Isaac's replies give unique access to his current thinking on a wide range of topics.

I speak for all the authors when I wish Isaac many more productive years in philosophy. For a start, the criticisms and constructive proposals in this book should keep him busy for a while. Having said this, I hasten to add that I consider the eventuality that Isaac should run out of interesting new ideas when left to his own devices not seriously possible.

E.J.O.

Introduction

The Pragmatism of Isaac Levi

Erik J. Olsson

Isaac Levi's philosophical thinking has shown remarkable stability over the years. Basically, it all started with his first book, *Gambling with Truth*, which outlines a research program whose key element is the decision-theoretic reconstruction of epistemology. Much of the rest of his work in epistemology has been devoted to extending and implementing this original program. With one important exception, there is little in his philosophical picture that has changed radically over the years. There have been changes, to be sure, but they have taken place at the level of detail rather than at the level of fundamental principle. The main exception is the issue of fallibilism. Starting out as a fallibilist, Levi became an infallibilist in the 1970s. The problem is that the corrigibility of our view suggests its fallibility: If we agree, as we must, that our view may change in the future, then it seems that we are never entitled to accept as true any claims of empirical substance now. But we do accept things as true now. Levi writes, in retrospect, that in the 1960s he unwittingly solved this problem for himself "by remaining in a fog of confusion" (Levi 1984, p. xiv), adding that by 1971 he had reached the conclusion that corrigibility and fallibility are best kept separate and, in particular, that endorsing corrigibilism is compatible with rejecting fallibilism. The paper "Truth, Fallibility and the Growth of Knowledge" was the first expression of this important revision. It was accepted for publication in 1975 but not actually published until 1983. The paper was reprinted in Levi (1984).

Levi is a truly systematic philosopher. The purpose of the following text accordingly is to describe his position in a way that reveals its internal coherence. I intend to do so without diving too deeply into the technical details. I want to show how arguably most of Levi's work in epistemology rests on four cornerstones: the belief-doubt model, the injunction against roadblocks in the path of inquiry, the unity of reason thesis, and the commitment-performance distinction. The first three elements undoubtedly belong to the tradition of

American pragmatism. The commitment-performance distinction may have some support in Dewey's work. In any case, Levi's epistemological thinking cannot be appreciated fully unless these cornerstones of his pragmatism are kept firmly in mind. This way of describing Levi's pragmatism departs somewhat from how Levi himself usually explains it, and I hope that it will prove useful as providing an alternative perspective from which to approach his epistemological work. The second purpose is to provide a conceptual map of the chapters in this book. I will try to indicate how the themes they address touch on central issues in Levi's philosophical thinking.

1. THE BELIEF-DOUBT MODEL

According to the Cartesian tradition in epistemology, we should start out in epistemology by doubting everything that can coherently be questioned. These efforts will, it is maintained, lead to a point where doubt is no longer possible, to a solid foundation in which our further beliefs can somehow be grounded. The recommendation to engage in methodological doubt is characteristically combined with an account of the latter according to which the mere logical possibility of error is sufficient to render a claim doubtful. The epistemological task, then, is taken to be one of recovering as many of our old beliefs as possible from a foundation that contains little more than logical trivialities.

Charles Sanders Peirce famously rejected the Cartesian epistemological picture, insisting that coherent doubtfulness is a much more exclusive property than Descartes would have us believe. All our beliefs, insofar as they are genuine convictions of ours, are things we accept as true without a moment's hesitation. There is, on our part, no "real and living doubt" that they are true. It is of course logically possible that a given empirical belief of ours is false. But this, Peirce thought, is beside the point. The mere logical possibility of error does not render a claim genuinely doubtful. By the same token, putting down a sentence in the interrogative form does not occasion real as opposed to "paper" doubt. We need positive reasons to doubt. Usually, doubt is occasioned by surprising experience.

Underlying Peirce's criticism of Cartesian epistemology is his belief-doubt model, according to which belief is an idle state that is satisfactory as it is. There is no point in inquiring further because there is no serious possibility that things are otherwise than we believe them to be. The matter is already settled. The intellectually pleasant state of belief can be disrupted, usually by the occurrence of an unexpected event of some sort, in which case the inquirer enters a state of doubt. The latter is a disharmonious state that the inquirer tries to avoid by engaging in inquiry, the goal of which is at least partly the fixation of a new belief.

The belief-doubt model is a central component of Isaac Levi's pragmatist epistemology. Levi insists, as Peirce did before him, that the beliefs we already entertain are in no need of justification. That is to say, there is no need for a person to justify his full beliefs *to himself*. According to the Cartesian line of thought, by contrast, a person's current beliefs do need to be justified even to that person himself.

To take an example from Levi's latest book, *Mild Contraction*: Before the invasion of Iraq, Bush, Chaney, and Rumsfeld were presumably not in doubt as to whether Saddam was in possession of weapons of mass destruction (WMDs). This matter was considered settled already. There was no point in letting the weapons inspectors continue their mission because there was no serious possibility that he would lack such weapons. If the Peircean belief-doubt model is correct, Bush and his associates did not at that point have to justify their belief in the existence of WMDs to themselves.

It is of course compatible with Peirce's view that a person may be in a situation that calls for her to justify her belief *to others*. While Bush and his associates did not at the time have to justify their belief in Saddam's possession of WMDs to themselves, they arguably had to justify it to the general public and to the U.N. We would not contradict Peirce if we were to claim in addition that a person should be able to justify any decision to *change* her convictions – even to herself. Many of those who initially believed fully that Saddam had WMDs believe now, with hindsight, that he did not have any after all. From what I have heard, we can count Bush and his associates to this lot. Be that as it may. At one point these people changed their convictions. This change, like any other, can itself be subject to justification. From this point of view, one of the major challenges facing epistemology is to spell out the conditions under which a given change in view is justified.

Cartesian epistemology is closely related to what Levi calls pedigree epistemology. Whereas Cartesian epistemology is first and foremost occupied with the question of what one can coherently and legitimately doubt, pedigree epistemology more directly concerns the nature of knowledge. What is common to pedigree epistemologists is that they are in a sense backward-looking. Roughly speaking, they focus not on how a given belief could be useful in the future but rather on how that belief was acquired in the past (Levi 1980, p. 1). As I understand Levi, the majority of contemporary epistemological theories qualify as pedigree epistemologies. Reliabilists, for example, insist that a belief qualifies as knowledge only if it was reliably acquired. Foundationalists, on the other hand, require that beliefs should be traceable to impeccable first principles. What reliabilists and foundationalists have in common is their preoccupation with pedigree of one sort or the other. Levi, by contrast, proposes that "[e]pistemologists ought to care for the improvement of knowledge rather than its pedigree" (ibid.).

In chapter 12 below, I question Levi's reasons for rejecting all forms of pedigree epistemology. Once the social aspect of knowledge is taken into account, a concern with pedigree is perfectly in order, or so I argue. The missing social dimension of knowledge is explored from a slightly different perspective in chapter 3. Philip Kitcher argues that Levi's approach needs to be extended to recognize the intricate ways in which social factors affect the modification of our beliefs.

As Levi has observed, the belief-doubt model has far-reaching consequences for the regulative role of truth in inquiry. It is commonly believed that what an inquirer should strive for, at a given point in her inquiries, is to arrive at the true, complete theory of the world, or at least of the relevant part of it that she takes interest in. According to the belief-doubt model, the inquirer is absolutely sure at that point that her current beliefs are true. This means that from her perspective, the true, complete theory of the world must form a superset of her current beliefs. It follows that if the inquirer gives up anything currently fully believed, she incurs a risk that she will not restore it at any point further down the line of inquiry. Indeed, for all she knows, she may even end up accepting its negation. In either case, she would undermine the effort to converge on the true, complete theory. Hence the only kind of belief revision she can justifiably engage in is expansion – the mere addition of new beliefs. But this is absurd, for it means that we can never come to doubt what we once believed to be true. Our beliefs become incorrigible.

Since Peirce subscribed to this view concerning the regulative role of truth – which Levi calls "messianic realism" – to the corrigibility of our view and also, of course, to the belief-doubt model, there is a serious conflict in his doctrine. Levi has sought to avoid trouble by rejecting Peirce's messianic realism. According to Levi's own "secular realism," what an inquirer should strive for at a given point in her inquiries is merely to obtain new true (error-free) information at the next step of inquiry: "inquirers should be concerned to avoid error as judged by the current doctrine only for changes of the current doctrine and not for any subsequent changes" (Levi 1998, p. 198). Thus, as for avoidance of error it is of no concern to the inquirer what happens further down the line of inquiry. In particular, it is of no concern to her whether or not a proposition once believed to be true will later fail to be believed or even be denied. Yet – and this turns out to be crucial – when we assess informational value, as opposed to avoidance of error, we do have the option of looking further down the line of inquiry:

> This view [secular realism] does, indeed, undermine the idea of scientific progress as progress toward the truth. It need not, however, undermine all conceptions of scientific progress. Inquiry does not get off the ground without demands for information, programs for research that aim, among other things, to obtain more comprehensive and informationally more valuable doctrines. Our goals in seeking more information are, on the view I have been advancing, far from myoptic. We do look ahead many steps down the line. There is nothing in secular realism that mandates myopia with respect to demands for valuable information but only with respect to avoidance of error. (Levi 1991, p. 163)

Hence we may, at a given point, anticipate that a very comprehensive and informationally valuable state of full belief can best be reached by first contracting parts of our present doctrine so as to make room for subsequent improvements. Secular realism, as opposed to the messianic variety, can be combined with a commitment to the corrigibility of our view.

Levi's interpretation of Peirce's belief-doubt model receives scrutiny in Cheryl Misak's contribution to this volume, chapter 1. She argues that the gulf between Levi's position and Peirce's is not as wide as Levi takes it to be. Chapter 2, by André Fuhrmann, focuses on Levi's critical view of truth in the limit and its place in inquiry. In Fuhrmann's view, absolute truth does play a role in inquiry, viz., to provide a reason for changing one's theoretical preferences.

2. THE INJUNCTION AGAINST ROADBLOCKS IN THE PATH OF INQUIRY

Another pragmatist component of Levi's thought that contributes to its distinctive character is the injunction against placing roadblocks in the path of inquiry. This notion, too, derives from Peirce, who thought that his principle "deserves to be inscribed upon every wall of the city of philosophy" (Peirce 1955, p. 54). Peirce eloquently defends his principle in the following passage:

> Although it is better to be methodological in our investigations, and to consider the economics of research, yet there is no positive sin against logic in trying any theory which may come into our heads, so long as it is adopted in such a sense as to permit the investigation to go on unimpeded and undiscouraged. On the other hand, to set up a philosophy which barricades the road of further advance toward the truth is the one unpardonable offence in reasoning, as it is also the one to which metaphysicians have in all ages shown themselves the most addicted. (Ibid.)

It is interesting to study Peirce's list of possible offenses against his principle. The first is to claim absolute certainty of matters of fact. The history of science reveals that many theories that were once taken to be the absolute truth later proved to be plainly false. Therefore, Peirce reasons, we should refrain from making absolute assertions now. As Levi has observed, Peirce's fallibilism – of which the argument just given is an expression – is in conflict with his belief-doubt model. For it follows from the latter that we must judge our current beliefs to be absolutely true. Moreover, pessimistic induction from the history of science is, on closer scrutiny, incoherent:

> Keep in mind that the judgment that the past record of inquiry is strewn with error (as well as truth) is predicated on the assumption that the current perspective is error free. For if the current perspective is not error free, on what basis do we judge the past record to be strewn with error? How can we judge that the current doctrine contains error by appealing to the premise contained in the current doctrine that false beliefs appear in past inquiry? What principle of selectivity entitles us to judge this element of the current doctrine true and insist that the rest contains error? Surely we ought to respect the total (relevant) evidence requirement. Here the total evidence is constituted by the current doctrine. (Levi 1998, p. 197)

Levi has sought to combine infallibilism (the rejection of fallibilism) with corrigibilism: We can be absolutely sure that our current beliefs are true and yet acknowledge that new evidence may be forthcoming that would make us change our view. Clearly, once we acknowledge the corrigibility of our view, absolute certainty is no obstacle in the path of inquiry. But for Levi's position to be convincing, it needs to be shown that one can theorize sensibly about belief correction and, above all, belief contraction. Hence, devising a convincing theory of belief contraction becomes an urgent project to which Levi has contributed in a number of papers and books. For detailed accounts, see Levi (1991, 1996). *Mild Contraction* is entirely devoted to this topic.

The second offense is one that Levi presumably could subscribe to without any qualifications. It lies in maintaining that this or that can never be known. Peirce's compelling example concerns Auguste Comte's contention that mankind would forever remain deprived of knowledge of the chemical composition of the fixed stars. But, as Peirce goes on to remark, "the ink was scarcely dry upon the printed page before the spectroscope was discovered and that which he had deemed absolutely unknowable was well on the way of getting ascertained" (Peirce 1955, p. 55). One should not make risky assertions about what may or may not be known in the future. Clearly, this is but an aspect of the corrigibility of our beliefs to which Levi has always been firmly committed.

The third "philosophical stratagem for cutting off inquiry" (ibid., p. 55) consists in maintaining that there are fundamental facts that are utterly inexplicable because there is nothing beneath them to know. Against this strategy, Peirce holds that it is no explanation of a fact to pronounce it as inexplicable and also that no reasoning could ever justify such a conclusion. Finally, we should not, in Peirce's view, hold that a law or truth has found its last and perfect formulation "and especially that the ordinary and usual course of nature never can be broken through" (ibid., p. 56). In practical terms this means that one should never engage in "absolute denial of an unusual phenomenon" (ibid.). This point is closely related to his first contention about absolute assertion and is for similar reasons not obviously correct.

Levi has put Peirce's no roadblocks principle to intriguing new uses. First and foremost, it serves to motivate the structure of what he calls conceptual frameworks. An inquirer's conceptual framework at a given time is the class of all states of full belief that are, in some weak sense, available for the inquirer at that time. If K_1 and K_2 are such potential states of full beliefs, then, Levi maintains, their join is also a potential state of full belief. The join consists of exactly those things that K_1 and K_2 have in common. Being in a belief state corresponding to the join of K_1 and K_2 is suspending judgment between these two states. The existence of a potential belief state representing the join is justified as follows:

> [C]onsider two inquirers, X and Y, sharing a common framework. X is in state K_x and Y is in state K_y. On some occasions it may be desirable for both X and Y to modify their views by adopting a belief state representing the shared agreement or common ground between them. To do this entails that they both give up informational value and, hence, incur a cost that they seek to minimize. In particular, they do not want to give up any more information than will be needed to bring them into agreement. The assumption of the existence of the join of K_x and K_y allows for the conceptual availability of such a move to both X and Y. It does not claim that exercising the option is always or even sometimes justifiable. However, we should not preclude such moves at the outset by denying that belief states representing such shared agreements are conceptually available. To do that violates the Peircean injunction against placing roadblocks in the path of inquiry. (Levi 1991, p. 13)

The existence of the meet of two potential states of full belief is similarly justified with reference to Peirce's no roadblocks principle.

The notion that it is always possible to suspend judgment between two states of full belief or theories is central in Levi's criticism of the incommensurability thesis of Kuhn and Feyerabend. As Levi interprets these authors, they hold that conflicting theories may be incommensurable in the sense that there is no common ground from the point of view of which their relative merits could be neutrally assessed. This is a view that Levi rejects:

> This join of K_1 and K_2 is the state of suspense that is the common ground to which X in state K_1 and Y in state K_2 could move if they were concerned to engage in joint inquiry that begged no questions against the other's point of view. To deny the availability of such a potential state of full beliefs (as authors writing in the tradition of Feyerabend and Kuhn often do) is to place roadblocks in the path of inquiry. Pragmatists will condone this practice only in the face of an impossibility theorem.
>
> (Levi 2002, p. 214)

Similarly, Levi wrote in an earlier work that "[t]o rule out in advance of inquiry the possibility of resolution by insisting that it involves a choice between incommensurables is to place roadblocks in the path of inquiry" (Levi 1984, p. 141).

In addition, the principle that suspension of judgment is always an option underlies Levi's view that certain evaluations of hypotheses lack truth value. This goes, in particular, for appraisals of truth value–bearing hypotheses with respect to subjective probability:

> Suppose to the contrary that such appraisal has truth value. That is to say, if X assigns h degree of credence r, he fully believes that h is objectively probable (in some sense) to degree r.
>
> Consider now a situation where X suspends judgments as to whether the degree of objective probability that h is 0.4 or 0.6. Let y be his degree of credence that the objective probability is 0.4 and $1 - y$ that the objective probability is 0.6. X's degree of credence that h is, under these circumstances, equal to $0.4y + (1 - y)0.6$. As long as y is positive and less than 1, X's degree of credence that h must be different from 0.4 and from 0.6. But this means that X must fully believe that the degree of credence is different from 0.4 and from 0.6 counter to the assumption that he is in suspense between these two rivals. (Levi 1984, pp. 156–7)

The common structure of Levi's striking arguments that this or that evaluation lacks truth value is that, if they have truth value, genuine suspension of judgment is not possible. But this runs counter to Peirce's injunction against obstructing inquiry. Hence, such evaluations lack truth value.

Levi has sought to motivate an alternative view of doubt that applies to attitudes that lack truth values. As for probability judgment, he has argued that we should be prepared to adopt credal states of hypotheses that are indeterminate and that allow many diverse distributions to be permissible. The set of permissible distributions should be a convex set: It should contain any linear combination of distributions in the set. Chapter 7, by Henry E. Kyburg, investigates the relation between convexity and conditionalization where the latter is taken as a principle for how to update one's probabilities in the face of new evidence. In chapter 6, Nils-Eric Sahlin defends Levi's way of representing probabilistic ignorance as part of a Socratic approach to decision making whereby experts can gain trust by admitting and communicating uncertainty. Probability is also the topic of chapter 8, by D. H. Mellor. The problem here is how knowledge of chances determines probability judgments to be used in practical deliberation and scientific inquiry. Mellor argues that Levi's view on this subject differs less than he thinks from its rivals. Chapter 9, by Wolfgang Spohn, is concerned more generally with how to conceptualize degree of belief and the relation between graded and absolute belief. Rejecting a probabilistic rendering of degree of belief, Spohn proposes an account in terms of so-called ranking functions. His chapter is devoted to spelling out the main differences between this approach and Levi's.

Evaluations of hypotheses with respect to serious possibility are also said to lack truth value. This claim plays a pivotal role in Levi's highly original criticism of modal realism. Levi proposes that the relevant notion of possibility is that of serious possibility. This notion is subject-relative. A proposition *p* is seriously possible for a subject (at a given time) if and only if *p* is consistent with her full beliefs (at that time). In relating the notion of possibility to an inquiring subject, Levi is making it potentially important in theoretical inquiry and practical deliberation. Now let us grant that evaluations of hypotheses with respect to serious possibility lack truth values. It would follow that counterfactual conditionals construed as hypothetical appraisals with respect to serious possibility lack truth value as well. In Levi's view, counterfactuals have truth values only to the extent that they are construed not as evaluations but as descriptions of the agent's conditional evaluations with respect to serious possibility. Against this background, the problem with Jaakko Hintikka's and David Lewis's approaches is that "[b]oth views imply that subjunctive conditionals have truth values and, moreover, that the truth conditions make no reference to the subjective state of the utterers (except, of course, insofar as such subjective states are described in the antecedents of consequents of such conditionals)" (Levi 1984, p. 157). Either these theorists grant that the relevant notion of possibility is serious possibility, in which case they cannot assign truth values to counterfactuals in the way they do, or they deny that the relevant notion of possibility is serious possibility, in which case it is unclear how their theories can be of any relevance to inquiry and deliberation:

I cannot prove conclusively that realistically construed notion of de dicto modality both conditional and categorical are *verdoppelte Metaphysik*. But the onus is on those

> who deny this to explain why the introduction of such conceptions is not gratuitous insofar as we are concerned with questions pertaining to epistemology, scientific inquiry and practical deliberation. (Ibid.)

Finally, Levi has applied Peirce's injunction against roadblocks in his discussion of the so-called rationality assumptions built into the conditions that entail Arrow's impossibility theorem. Here, too, the admissibility of suspending judgment plays a key role in the argumentation. For the details, see Levi (1984, pp. 247–70).

Who would have thought initially that Peirce's injunction against roadblocks should have had repercussions for the interpretation of counterfactual conditionals or Arrow's theorem? Levi should be credited with exploring the consequences of some pragmatist principles in greater depth and more consistently than others have done before him. Thanks to him, we are in a better position to appreciate the perhaps surprising force of the pragmatist tradition of thought.

In chapter 4, Bengt Hansson argues that while fallibility and corrigibility are indeed independent notions, infallible items of knowledge should not be identified with those that are maximally certain. Another issue of interest in this connection concerns inconsistency. Levi thinks that the inconsistent state of full belief should be part of every conceptual framework. Once inconsistent, our state of full belief cannot function properly as a standard of serious possibility. There is nothing in terms of which we can distinguish those possibilities that are serious from those that are not. Our standard of serious possibility has broken down. A problem that Levi until just recently had not given the attention it deserves is that there seems to be no rational way to escape from inconsistency. For any such way would have to be based on the current state of full belief. But if the current state is inconsistent, it is useless for purposes of inquiry and deliberation. In particular, it is useless for inquiry into how to get rid of the inconsistency. Inconsistency, then, is the ultimate roadblock of inquiry. It is "epistemic hell," to use a phrase coined by Peter Gärdenfors (1988). In response to this sort of criticism, as leveled by myself (in Olsson 2003), Levi has recently changed his theory of contraction. His new position is that contraction from inconsistency must be construed not as a matter of deliberation but as a matter of routine (Levi 2003, 2004b). An alternative strategy is explored in Otávio Bueno's contribution, chapter 5. Bueno argues that Levi's position could be strengthened by making room for inconsistency in a way that, he believes, does not jeopardize any commitment to pragmatism.

3. THE UNITY OF PRACTICAL AND THEORETICAL REASONING

The third cornerstone of Levi's pragmatism concerns the connection between practical deliberation and theoretical inquiry. Levi is committed to practical and theoretical reasoning being in a deep sense one and the same. As we will see, his position contains several distinctive components that should accordingly be given separate attention.

According to popular opinion the distinctive mark of scientific as opposed to practical matters is that science is value-free or value-neutral. Against this, Levi maintains that science inquiries are just as value-laden as are practical investigations. Scientists qua scientists must make value judgments. The difference is that the values that should be promoted in scientific inquiries are different from, and irreducible to, the practical values that figure in political, economic, moral, or aesthetic deliberations: "The reconstructed version of value-neutrality that I favor denies this reductionist view and insists that scientific inquiries seek or ought to seek to promote values and goals distinctive of the scientific enterprise" (Levi 1984, p. ix). As scientific or cognitive values Levi counts logical strength, simplicity, explanatory power, and the like. They are all subsumable under the umbrella concept of informational value.

Does the autonomy of scientific values create a questionable dualism between theory and practice? Levi's answer is in the negative. First, both kinds of activity are goal-driven. Just as proposals for how to act should be evaluated in terms of efficiency in realizing given practical ends, so too proposals for how to change one's theory should be evaluated in terms of efficiency in realizing given cognitive ends.

At one point Levi goes as far as claiming that the goal-driven nature of theoretical rationality makes such rationality but a species of practical rationality:

> [T]he classical pragmatists . . . certainly were in favor of an integrated understanding of practical and theoretical rationality. Science differs from what Dewey called "common sense" in its goals. As a consequence, it also differs in its methods. But insofar as rationality plays a role, it is means-end rationality in both cases. That is to say, it is practical rationality. (Levi 2004a, p. 244)

In Dewey's terminology, common-sense deliberation focuses on practical issues. Up to this point, "practical rationality" has referred to the sort of rationality that aims at the choice of a practical action, whereas "theoretical rationality" has referred, roughly, to the sort of rationality that aims at the fixation of belief. Yet in the passage just quoted, Levi is using the term “practical rationality” in the new sense of “goal-driven rationality.” To call all goal-driven rationality “practical” does not add anything, except conceptual confusion, to the point already made that both rationality aiming at the choice of practical action and rationality aiming at the settlement of opinion are goal-driven activities.

At any rate, Levi does not merely want to suggest that practical deliberation and theoretical inquiry are similar in the sense that both are goal-driven; he also proposes that the same principles are at work in both cases: “the principles of rational choice or rational goal attainment governing deliberation in science ought to be the same as those regulating the rational attainment of moral, political, economic, and other practical objectives” (Levi 1980, pp. 71–2). Hence, “[t]he difference between theoretical inquiry and practical deliberation is a difference in goals and not a difference in the criteria for rational choice that regulate efforts to realize these goals” (Levi 1984, p. 72). What Levi is suggesting is that there is a far-reaching *structural* unity between practical and theoretical inquiries. In his own words, “[w]hat is ‘pragmatic’ about pragmatism is the recognition of a common structure to practical deliberation and cognitive inquiry in spite of the diversity of aims and values that may be promoted in diverse deliberations and inquiries” (Levi 1991, p. 78).

In concrete terms, the structural unity thesis suggests that it may be worthwhile to apply Bayesian decision theory not only in the practical realm, but also in the cognitive domain. Levi has made substantial contributions to this area of research since the 1960s. His elegant Bayesian account of inductive acceptance, first formulated in *Gambling with Truth* and slightly modified in “Information and Inference” (Levi 1984, pp. 51–69), is a milestone on this path of philosophical inquiry. This account has interesting implications for the lottery paradox and other long-standing problems of induction. In *Fixation* and later works, he has addressed the problem of reconstructing belief contraction, too, as a problem of rational choice. The decision-theoretic perspective has substantial consequences for contraction as well. It entails, for example, that if the inquirer in belief state K retracts a belief by entering a new weaker belief state K', then contracting to K' must have been one of the inquirer’s options while in K. Several other theories of contraction, among them the celebrated partial meet approach of Alchourrón, Gärdenfors, and Makinson (1985), fail to comply with this simple rule.

An inquirer’s standard of serious possibility is constituted by his full beliefs. Everything that is compatible with the inquirer’s full beliefs is judged seriously possible from his point of view. A further aspect of Levi’s unity of reason thesis is the claim that the standard of serious possibility used in cognitive inquiries should be the same as that used in practical deliberations: "rational X should, during any minimal interval of time, be committed to a single standard for serious possibility both for theoretical inquiry and for practical deliberation" (Levi 1980, p. 16). An inquirer should refrain from using one set of background assumptions or full beliefs in theoretical matters and another set in practical matters. A double standard of serious possibility should be avoided.

As noted in Levi (1980, pp. 16–17), Peirce seems to have taken the opposite view:

> If a proposition is to be applied to action, it has to be embraced or believed without reservation. There is no room for doubt, which can only paralyze action. But the scientific spirit requires a man to be at all times ready to dump his whole cartload of beliefs, the moment experience is against them. The desire to learn forbids him to be perfectly cocksure that he knows already. . . . Thus the real character of science is destroyed as soon as it is made an adjunct to conduct; and especially all progress in the inductive sciences is brought to a standstill. (Peirce 1955, pp. 46–7)

Apparently, Peirce is here implying that in practical deliberation some logical possibilities are discounted as not being seriously possible. He is also saying that in scientific inquiry all logical possibilities are serious. It follows that the standard for serious possibility used in practical deliberation cannot coincide with that employed in theoretical inquiry. Yet, this reading of Peirce as advocating double standards of serious possibility can be questioned, and Levi has recently stated (in personal communication) that he now finds this interpretation implausible given the context in which the remark was made. For more on this, see Cheryl Misak's discussion of the gap between science and vital matters in chapter 1. In chapter 15, Mark Kaplan argues that whether a person knows that something is the case can be affected by what is practically at stake if she acts on her belief in her current circumstances. By raising the stakes, one can apparently make one's knowledge go away. This, if correct, would shed doubt on Levi's position regarding the autonomy of theoretical reasoning vis-à-vis practical reasoning.

Two authors write on the topic of abduction, which Levi, again following Peirce, conceives of as the initial stage in theoretical inquiry and practical deliberation at which the alternative answers to the inquirer's question are identified. In chapter 10, Maurice Pagnucco provides an overview of Levi's theory of abduction, comparing it with other accounts, primarily those that have been devised by researchers in artificial intelligence. Levi's conception turns out to be quite distinct from those other accounts. In chapter 11, I object to Levi's proposal to view all alternative hypotheses of relevance to a given question as being also potential answers to that question. Roughly speaking, hypotheses of a disjunctive nature, while being of relevance in inquiry, are not in any interesting sense potential answers.

Several authors have chosen to comment on Levi's decision theoretic account of epistemology. In chapter 12, I argue that while Levi's solution works very well considered in isolation, combining it with his individualistic conception of knowledge leads to the uncomfortable result that one can know that one's lottery ticket will not win. Hans Rott's chapter (13) contains a concise summary of Levi's account of belief expansion and contraction together with some penetrating criticisms. Levi's theory of contraction and sequential change is investigated at length in chapter 14, by Horacio Arló Costa, which aims at mending bridges between the decision theoretic perspective and other contemporary work in belief change done mostly by computer scientists. It is worth noting that both Rott and Arló Costa base their commentaries on Levi's most recent book, *Mild Contraction*, which contains an account of contraction that differs in important respects from his earlier theory.

4. THE COMMITMENT–PERFORMANCE DISTINCTION

While the classical pragmatists differed as to what inquirers should strive for more precisely, they arguably shared the view that they need to justify changes in view by showing that one change is better than the alternatives for the purposes of promoting the goals of the given inquiry. An activity cannot be goal-driven unless the agent is capable of adjusting her behavior so as to promote the goal she is trying to attain. Changes in view should be no exception to this rule. Suppose, however, that beliefs are merely dispositions to linguistic and other bodily behavior, as many respectable philosophers have indeed argued. Then the question of how to justify changes in view does not even arise, as the inquirer lacks the control necessary to be held accountable for such changes. If, as Quine and others maintain, coming to believe is merely a matter of responding in a certain ways to external stimulation, pragmatism is seriously in error.

Levi's solution to this problem is one of his most original contributions to American pragmatism. His proposal is that we distinguish between changes in commitment and changes in performance. As Levi reads Dewey, the latter was primarily interested in changes in commitment. Such changes can plausibly be subject to the agent's direct control and they are therefore the sort of thing one would typically need to justify. In the case of beliefs, the commitments of relevance are doxastic commitments, that is, commitments to believe something. A commitment to believe can be seen as a promise to believe (although there are certain important differences as well; see Levi 2002, p. 228).

Changing a doxastic commitment is one thing, implementing it quite another. The agent may fail in his performance to live up to his commitments. He may fail to believe what he is committed to believe. This can happen for many different reasons. Levi mentions, as possible causes, lack of calculating capacity, subjection to an emotional storm, or distraction from self-critical reflection.

To take an example, a person may believe initially that the French city of Nancy is south of Hamburg without entertaining any particular view about the location of Helsinki. On consulting a map she comes to believe that Helsinki is north of Hamburg. That would count as a change in doxastic commitment. A new commitment about the location of Helsinki has been added to her old stock of commitments. Her two geographical commitments together entail the further commitment to believe that Nancy is south of Helsinki. Suppose, however, that the person does not at first realize that she is committed to believing that Nancy is south of Helsinki. At a later point she realizes this. This change would count as a mere change in performance. In believing that Nancy is south of Helsinki, she is closer than she was before to fulfilling all her doxastic commitments.

Levi's theory of belief revision is a theory of commitment change, not a theory of performance change. The states of full belief in his theory are ideal states in which all doxastic commitments are realized. They should be seen as equilibrium states similar to the objects studied by classical thermodynamics and economics. The theory describes how changes take place from one equilibrium state to another. It is silent about how to go from a nonequilibrium state to an equilibrium state. This problem is deferred to the separate study of performance change.

Nevertheless, I fail to see how invoking the commitment-performance distinction could serve to neutralize the antivoluntaristic objection that was raised at the beginning of this section. A theory of commitment change is empty unless taking on a commitment to believe is at least positively relevant to actually implementing the commitment. If there were no such relation between commitment and performance, Levi's theory would lack all significance for actual inquiry. This would mean that Levi's reply to Quine and other dispositionalists is partly question-begging, as they would presumably reject the notion that deciding to believe is positively relevant to actually believing.

Still, for those who are already sympathetic to this notion, the commitment-performance distinction does make a lot of sense.

In fairness to Levi it should be mentioned that voluntarism is not the only issue he intends to tackle with his distinction. He also believes that his theory has the virtue of reducing two mysteries for naturalism to one. The two mysteries are the obstacles to naturalism presented by the naturalistic fallacy and the gap between nature and meaning. By taking attitudes in general to be commitments, there is, he submits, hope that the question of meaning can be understood as a question about values. Another advantage he sees in this manner of theorizing is that it enables us to give up "the pretence that principles of rationality are primarily used for the purpose of explanation and prediction" (Levi 2002, p. 223). There is much more to be said about this than the space allocated to this introduction allows. The interested reader should consult Levi's 1997 book *The Covenant of Reason*. Levi (2002) is a good summary of Levi's view on rationality and the commitment–performance distinction.

Levi's preoccupation with commitment rather than performance represents one sense in which his epistemological picture abstracts from the vagaries of actual human inquiry. Chapter 16, by Sven Ove Hansson, seeks to identify the different idealizations that are involved in Levi's theorizing. According to Levi, principles of rationality – be they theoretical or practical – serve two different purposes: They regulate changes in commitment and performance. They provide criteria by means of which changes in commitment can be evaluated, and they indicate the standards to which our performance should conform (Levi 1997, p. 16). Several chapters address Levi's theory of rationality. A pragmatic argument for a principle is an argument that appeals to the desirable/undesirable consequences of that principle's satisfaction/violation. Chapter 19, by Wlodek Rabinowicz, focuses on pragmatic arguments for various rationality constraints on beliefs and preferences, and on Levi's view of the status of such argument. In his contribution, chapter 17, Wolfram Hinzen confronts Levi's view on rationality with another, more naturalistic account. Naturalism and commitment are also central themes in Akeel Bilgrami's chapter (18), in which these issues are discussed in the context of psychoanalytic theory. Bilgrami argues, among other things, that the concept of a neurosis is inherently normative, involving a failure of one's dispositions to accord with one's commitments. The general nature of dispositions is the central issue in chapter 20, Johannes Persson's contribution, which compares Levi's view – as first stated in an early joint paper with Sydney Morgenbesser (1964) – with that of Jon Elster.

REFERENCES

Alchourrón, Carlos, Peter Gärdenfors, and David Makinson. 1985. "On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision." *Journal of Symbolic Logic* 50: 510–30.

Gärdenfors, Peter. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, Mass.: MIT Press.

Levi, Isaac. 1967. *Gambling with Truth: An Essay on Induction and the Aims of Science*. Cambridge, Mass.: MIT Press.

Levi, Isaac. 1980. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. Cambridge, Mass.: MIT Press.

Levi, Isaac. 1983. "Truth, Fallibility and the Growth of Knowledge." In R. S. Cohen and M. W. Wartofsky (eds.), *Language, Logic and Method*, pp. 153–74. Dordrecht: Reidel.

Levi, Isaac. 1984. *Decisions and Revisions: Philosophical Essays on Knowledge and Value*. Cambridge: Cambridge University Press.

Levi, Isaac. 1991. *The Fixation of Belief and Its Undoing: Changing Beliefs through Inquiry*. Cambridge: Cambridge University Press.

Levi, Isaac. 1996. *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inferences, and Nonmonotonic Reasoning*. Cambridge: Cambridge University Press.

Levi, Isaac. 1997. *The Covenant of Reason: Rationality and the Commitments of Thought*. Cambridge: Cambridge University Press.

Levi, Isaac. 1998. "Pragmatism and Change of View." *Canadian Journal of Philosophy* suppl. vol. 24: 177–201.

Levi, Isaac. 2002. "Commitment and Change of View." In José Luis Bermúdez and Alan Millar (eds.), *Reason and Nature: Essays in the Theory of Rationality*, pp. 209–32. Oxford: Oxford University Press.

Levi, Isaac. 2003. "Contracting from Epistemic Hell Is Routine." *Synthese* 135: 141–64.

Levi. Isaac. 2004a. "Corrigibility without Solidarity." In Elias L. Khalil (ed.), *Dewey, Pragmatism, and Economic Methodology*, pp. 240–51. London: Routledge.

Levi, Isaac. 2004b. *Mild Contraction: Evaluating Loss of Information Due to Loss of Belief*. Oxford: Oxford University Press.

Levi, Isaac, and Sydney Morgenbesser. 1964. "Belief and Disposition." *American Philosophical Quarterly* 1: 221–32.

Olsson, Erik J. 2003. "Avoiding Epistemic Hell: Levi on Pragmatism and Inconsistency." *Synthese* 135: 119–40.

Peirce, Charles Sanders. 1955. *Philosophical Writings of Peirce*, ed. Justus Buchler. New York: Dover.

1

Isaac Levi and His Pragmatist Lineage

Cheryl Misak

INTRODUCTION

Isaac Levi stands out as one of the most important philosophers who has worked in the pragmatist tradition. Like his predecessors, Charles Peirce, William James, and John Dewey, Levi insists that we must take practice and context seriously when we think about knowledge and truth. Each of the classical pragmatists followed through on this central insight in a different way and each has motivated a different kind of contemporary pragmatist. Levi's kind focusses on according our existing corpus of belief its actual and proper status in epistemology. What we are concerned with in inquiry – in seeking knowledge – is the *revision* of that corpus of belief as opposed to the *pedigree* or *origin* of belief. What we are concerned with is whether we should retain a commitment or whether we should abandon it in favor of an alternative commitment.

Levi seems to sometimes take himself to be closest to Dewey, in whose old department – Columbia – Levi spent the bulk of his career. They both focus on the problem-solving nature of knowledge. But it is more apt, I suggest, to think of Levi as the inheritor of Peirce's position. Levi has himself acknowledged the similarities. But he also identifies what he takes to be significant gulfs between his position and Peirce's. My aim in this chapter is to show that these are not as wide as they might first appear. I was fortunate to have the opportunity to study for a year with Levi and I have tried some of these arguments on him before.[1] Here I make one more attempt at persuasion.

PEIRCE ON THE FIXATION OF BELIEF

The *locus classicus* of Peirce's view of truth and knowledge is his 1877 paper 'The Fixation of Belief', in which he puts forward his 'doubt-belief' model

[1] My 'Peirce, Levi, and the Aims of Inquiry' grew out of a course paper written for Levi.

of inquiry. Inquiry begins with the irritation of doubt: It is prompted when a stable belief (an expectation) is upset by a surprising experience. Inquiry ends once we get another stable doubt-resistant belief – one that accounts for the surprising experience. If we were to have a belief that would always be immune to doubt – a belief that would forever fit with experience and argument – then that belief would be true. There would be nothing more we could ask of it. Since we can never know when a belief is like that, our beliefs are fallible: Any one of them might be shown by future inquiry to be false.

Fallibilism, however, does not entail that we ought to follow Descartes and try to bring into doubt all beliefs about which error is conceivable. Peirce, in 'The Fixation of Belief', ignites the pragmatist torch against Cartesian foundationalist epistemology – against thinking of justification in terms of global doubt. Were we to doubt a belief simply on the basis of a suspect pedigree, this doubt, Peirce argued, would be a 'paper' or 'tin' doubt – not the genuine article. Genuine doubt is always prompted when a particular belief or expectation is upset by something that surprises us. It is not prompted by a thought experiment that places all of our beliefs under a cloud of suspicion because their pedigree is not perfect. All of our beliefs are fallible, but they do not come into doubt all at once.

This view was never abandoned by Peirce. In 1905 we find him saying that

> there is but one state of mind from which you can 'set out', namely, the very state of mind in which you actually find yourself at the time you do 'set out' – a state in which you are laden with an immense mass of cognition already formed, of which you cannot divest yourself if you would. . . . Do you call it doubting to write down on a piece of paper that you doubt? If so, doubt has nothing to do with any serious business. . . . (CP 5.416[2])

Any philosophical discussion of knowledge must start with our current (Peirce often says 'commonsense') body of knowledge. From that starting point, knowledge is rebuilt in the serious business of inquiry bit by bit when experience forces inquirers to revise their beliefs. Inquiry 'is not standing upon the bedrock of fact. It is walking upon a bog, and can only say, this ground seems to hold for the present. Here I will stay till it begins to give way' (CP 5.589, 1898). Accepted hypotheses and theories ('established truths') are stable and believed to be true until they are upset by the surprise of experience. Only

[2] References to Peirce's *Collected Papers* are in standard form: volume number, paragraph number.

against such a ‘commonsense’ background of established truths can a belief be put into doubt and a new, better belief be adopted. The inquirer

> is under a compulsion to believe just what he does believe . . . as time goes on, the man’s belief usually changes in a manner which he cannot resist . . . this force which changes a man’s belief in spite of any effort of his may be, in all cases, called a *gain of experience*. (MS 1342, p. 2, undated)

Peirce links the scientific method to this epistemology. It is the method that pays close attention to the fact that beliefs fall to the surprise of recalcitrant experience. Hence it is a method that is suited to lead to the truth. On Peirce’s view, when we say that we aim at the truth, what we mean is that we aim at beliefs that would be forever stable. We aim at beliefs that would satisfy all our local aims in inquiry: empirical adequacy, coherence with other beliefs, simplicity, explanatory power, getting a reliable guide to action, fruitfulness for other research, greater understanding of others, and the like. Were a belief really to satisfy all of our local aims in inquiry, then that belief would be true. There is nothing over and above the fulfillment of those aims, nothing metaphysical, to which we aspire. It isn’t too much a simplification to say that a true belief is a belief that would forever account for experience and argument. (Peirce has an enormously broad conception of experience, on which argument is a kind of experience. You can be surprised by the force of an argument.[3]) Hence truth and the scientific method are especially suited to one another. For the scientific method is the method that tests beliefs by exposing them to experience and argument.

This epistemology and its accompanying view of truth are entirely general, despite the fact that Peirce allies them with the method of science. For what Peirce calls ‘science’ is extremely broad. Any inquiry that aims at getting a belief that would forever stand up to experience and argument abides by the method of science. He thought that metaphysics (when it is well conducted) and mathematics are part of science and are legitimate aspirants to truth. And so is ethics.[4]

LEVI ON THE FIXATION OF BELIEF

Levi is now the bearer of Peirce’s torch against Cartesian epistemology. He has echoed Peirce’s famous paper in a book of his own, titled *The Fixation of Belief and Its Undoing* (1991), and he returns to these themes in his new *Mild*

[3] See Misak 1991, pp. 21ff for elaboration.
[4] See Misak 2000 for a sustained argument.

Contraction.[5] Indeed, the title of that 1991 book reflects Peirce's argument better than Peirce's own title. We have seen that Peirce takes the undoing of belief – the fact that it is undone by recalcitrant experience and argument – to be what gives the scientific method is character and its special relationship with the truth.

Like Peirce, Levi turns his back on the basic requirement of Cartesian and indeed contemporary epistemology: the requirement that global doubt is the way to approach the task of justifying beliefs that are current and, hence, not subject to doubt. Peirce's notion of genuine or live doubt is transformed by Levi into the notion of serious possibility. An agent's corpus of belief is used as a resource in inquiry by serving as a standard by which to judge conjectures: as a standard for determining just what is and what is not a serious possibility. Not all logical possibilities are serious ones: Serious possibilities are those that are consistent with an agent's background body of belief. To serve as a standard of serious possibility, the beliefs in my corpus must be regarded by me as being true – I must not doubt the truth of my current beliefs nor think that they require justification. In *The Enterprise of Knowledge*, Levi labels this anti-Cartesian position 'epistemological infallibilism'.

This position does not entail that individuals should not be concerned with justification at all. The point is that the scope of the demand for justification is more modest than many epistemologists have thought. Justification is demanded only when a change in commitment or belief is demanded.[6]

The position also does not entail that inquirers should be sceptics who equate serious possibility with logical, mathematical, or conceptual possibility. We can fully believe a proposition even if it isn't a logical, mathematical, or conceptual truth.

So, on Levi's view, we are required to justify a belief only when we are in the business of revising it, only when belief is becoming unfixed. There are

[5] See also Levi 1980.

[6] With Peirce, Levi takes belief to be a commitment, although he does not go quite as far as Peirce in accepting Bain's definition of belief as a disposition to behave. For one thing, Levi thinks that if a belief involves a disposition to behave, the behaviors manifested in the disposition are a function not only of the agent's beliefs but of his values, desires, and preferences. For another, Levi takes belief not to be equivalent to having dispositions to manifest conviction or lack of conviction in a proposition. A belief is equivalent to having the commitment to have such dispositions. The extent to which I fulfill my doxastic commitments depends on my abilities and temperament – I may fully believe that p in the sense of having the doxastic commitment, yet I might fail to behave in a manner that fulfills the commitment. The commitment is there, whether I can live up to it or not. My abilities and temperment can of course be improved on: As Levi says, we spend fortunes on devices to enhance our memory, to improve our computational capacity, and to escape from the disabilities of depression and other emotional difficulties. See Levi's *Mild Contraction*.

three ways, he argues, to revise one's corpus of belief. The first is expansion, which can be inferential (deliberate) or routine. In the latter, one relies on a predetermined process to decide which hypotheses will be accepted: One can, for instance, rely on one's senses or on an expert. The second way to revise belief is contraction: An agent concludes that some belief in her corpus ought to be scrutinized and so she gives it up as an infallible truth. The last is replacement: changes of commitment from one hypothesis or theory to another. Much of Levi's reputation lies in the technical details of these methods of fixing and unfixing beliefs. And much of his work has involved grappling with a problem centering around replacement.[7]

MESSIANIC VERSUS SECULAR REALISM

Suppose I am (reasonably) concerned to avoid importing false beliefs into my corpus of belief. I must of course judge truth and falsity from my current point of view: My standard for serious possibility, truth, and falsity is my current corpus of belief. This means that if I am looking to justify the replacement of $\neg p$ by p, then I must start from my belief that $\neg p$. From that position, however, it has to seem that I will import false belief if I replace $\neg p$ with p. If I want to avoid error, I should regard such a move as unacceptable, and hence replacement is never going to be justified. It should be impossible, moreover, for a scientific revolution to occur. When a scientist 'converts' from one theory to another incommensurable theory, it seems that he deliberately imports beliefs he is initially convinced are false.

Levi suggests a two-step process as a way of avoiding this dilemma: I contract my corpus of belief by getting rid of $\neg p$. I don't import a false belief in taking this step. I will of course risk importing error subsequently. But when attempting to justify changing my doxastic commitments, I may be concerned to avoid error in implementing the very next step and not with avoiding error in subsequent steps. Levi calls those who seek to avoid error at just the next step 'secular realists'. The secular realist can sometimes justify replacement of $\neg p$ by p by justifying a sequence of contractions and expansions. The secular realist adopts as a goal of inquiry the acquisition of error-free belief. But the strategy for attaining the goal is myopic: He looks only at the first step, not at subsequent steps. And then in no step will he give up a truth for a falsehood. Once the contraction is implemented, he is no longer is certain that p is false. He may then, perhaps, find a justification for expansion, for adding p. The net effect is a justification of the replacement of $\neg p$ by p that

[7] See Levi 1980, 1991, and *Mild Contraction*.

is not direct and, hence, does not require indifference to the possibility of importing error at the next step. Choosing direct or one-step replacement would require just that.

'Visionary realists', on the other hand, not only are concerned to avoid error at the next step but are concerned to take into account the risk of importing false belief at later stages. 'Messianic realists' are in an even worse position: They hope to obtain at the 'End of Days' a system of full beliefs that is not only free of error but complete in the sense that all doubt about every issue is removed. Visionary and messianic realists will find the two-step strategy that Levi argues for unacceptable. For they will see that from their initial perspective, they will be exchanging a belief they take to be true for one that they take to be false.

Here is where Levi thinks he parts company with Peirce. Peirce, he charges, was a messianic realist and so cannot avail himself of the two-step strategy. Peirce is supposed to hold that the ultimate aim of inquiry is 'convergence on the true complete story of the world'(1980: 20).[8] Since rational agents should favour revisions that best promote the objectives they are commited to in inquiry, then from the Peircean's initial point of view, replacement is never rational. Replacing $\neg p$ with p (replacing a certainty with a certain falsehood) will frustrate the goal of coming closer to the one true complete story of the world.

Thus, Levi thinks that Peirce adopts an aim of inquiry that makes rational revision of belief impossible. He is stuck with the problem that Levi takes to dog every pragmatist: the problem of how to justify revision of belief if you take your current beliefs seriously. Only secular realists, from their point of view prior to contraction, can remove $\neg p$ from their corpus of belief with no risk of error.

I don't want to speak to the issue of whether secular realists, with their myopic strategy, can get away with their two-step dance. For whether they can or not, there is another route to avoiding the problem, a throughly Peircean route.

We first need to ask whether Peirce was really commited to the goal of reaching the complete truth in the long run. I think not.[9] The best way to interpret Peirce is to take him to be arguing that 'truth' is just a catch-all for all of those particular local aims of inquiry. When Peirce says that a true belief is a belief that would be agreed on in the long run, what he means is that were a belief or theory to satisfy all of our local aims in inquiry, then it would be

[8] Levi 1980, p. 20.
[9] See Misak 2004 and 2000 for a sustained argument.

true. As we specify our cognitive ends, we make more specific our concept of truth.[10] There is nothing over and above the fulfillment of those ends, nothing complete and nothing metaphysical, to which we aspire. The pragmatist steps away from metaphysically loaded accounts of truth and steps toward practice. A true belief is one that would be the upshot of our inquiries. Indeed, Peirce could not be clearer than he is in 'The Fixation of Belief':

> [T]he sole object of inquiry is the settlement of opinion. We may fancy that this is not enough for us, and that we seek, not merely an opinion, but a true opinion. But put this fancy to the test, and it proves groundless; for as soon as a firm belief is reached we are entirely satisfied, whether the belief be true or false. (CP 5.375)

And then:

> You only puzzle yourself by talking of this metaphysical 'truth' and metaphysical 'falsity' that you know nothing about. . . . If your terms 'truth' and 'falsity' are taken in such senses as to be definable in terms of doubt and belief and the course of experience . . . well and good: in that case you are only talking about doubt and belief. But if by truth and falsity you mean something not definable in terms of doubt and belief in any way, then you are talking of entities of whose existence Ockham's razor would clean shave off. (CP 5.416)

We must stay away from 'vagabond thoughts that tramp the public roads without any human habitation' (CP 8.112). Such thoughts about truth make it 'the subject of metaphysics exclusively'. We would do better to deflate such accounts of truth by linking truth to belief, assertion, experience, and inquiry.

The result of this linkage, Peirce argues, is that we should think of a true belief as a belief that would forever be assertible; a belief that would never lead to disappointment; a belief that would be 'indefeasible' or not defeated, were inquiry pursued as far as it could fruitfully go (CP 5.569, 6.485). This is very clearly not a view of truth that has it that our aim in inquiry is to get to the one complete True account of the world.

It is also very clearly a view on which a strategy exists to unfix belief and replace it with a new belief. From a believer's current perspective, $\neg p$ is taken to be true; it is not doubted. Then an experience or argument comes along that upsets the belief. $\neg p$ is then removed and inquiry is ignited for a replacement. If that replacement turns out to be p, so be it. In the first step, $\neg p$ is removed, for good reason. In the second, p is added, again for good reason. From the believer's initial perspective, what is important is the removal of $\neg p$,

[10] See Richardson 1994 for this way of putting the point.

as experience has spoken against $\neg p$. What replaces it is a matter for inquiry to determine. This is a strategy that is myopic in Levi's sense: It requires us to focus on the issue at hand.

THE GAP BETWEEN SCIENCE AND VITAL MATTERS

A second (related) mistake Levi sees in Peirce is harder to sort out and does indeed threaten to drive a wedge between the views of these two pragmatists. Peirce was frequently keen to insist that 'vital' or ethical matters[11] are not subject to the same problem-driven, experience-based, truth-aimed method as are scientific matters. He says that when trying to answer vital questions, we must eschew reason in favour of instinct, for we need to reach a definite conclusion quickly. Science, on the other hand, 'has nothing at stake on any temporal venture but is in pursuit of eternal verities . . . and looks upon this pursuit, not as the work of one man's life, but as that of generation after generation, indefinitely' (CP 5.589, 1898). Science, but not ethics, goes on the hope that 'the truth may be found, if not by any of the actual inquirers, yet ultimately by those who come after them and who shall make use of their results' (CP 7.54, 1902). Thus, science is concerned with truth and ethics is not. The flip side of the point, he suggests, is that 'really the word belief is out of place in the vocabulary of science' (CP 7.185, 1901). Science concerns itself with a 'formula reached in the existing state of scientific progress' – not a belief on which to act.

Peirce appears to offer us here an extreme kind of noncognitivism, where matters of ethics do not fall under the scope of truth, knowledge, and inquiry. In ethical matters, we do not aim at getting the answer in the long run, but, rather, we follow instinct, convention, and common sense in order to get an answer here and now. The preservation of the status quo seems inevitable. Indeed, Peirce is clear that this view, which he at times calls 'sentimentalism', 'implies conservatism' (CP 1.633, 1898).

Levi, with Dewey, wants to bring ethical problems under the scope of the problem-based pragmatist epistemology that seems so suited for it. So it looks as if Peirce and Levi do part ways. Indeed, because Peirce appears to offer us an extreme view of science on which the scientist has no commitments at all, they seem to part ways very radically. Everything I have said about Peirce thus far seems to be false.

[11] As Chris Hookway pointed out to me, a vital matter, for Peirce, is any urgent question about what we ought to do. The category of the vital is wider than the category of the ethical.

But once we understand what Peirce means by 'instinct', 'experience', and 'commonsense' and once we understand their roles in what he calls scientific inquiry, Peirce's distinction between science and ethics is not as odd as it first appears and it is not as opposed to Levi's position as it first appears. Instinct is part and parcel of science, and ethical matters are indeed suitable for scientific inquiry.

Peirce builds instinct into the very basics of the scientific method. He often characterizes that method as the method of abduction, deduction, induction. Abduction is a matter of coming up with (Peirce sometimes says 'guessing at') an explanation for a surprising experience. Once abduction has provided a hypothesis, one deduces consequences from it and then tests the hypothesis by induction.

Abduction provides science with new ideas, and because this guessing at new hypotheses is not codifiable, Peirce says that science advances by 'the spontaneous conjectures of instinctive reason' (CP 6.475, 1908, 5.604, 1901). When a surprising phenomenon needs explanation, instinct plays a central role. It provides the starting points of the scientific method – the hypotheses whose consequences are then tested. That is one way in which instinct, rather than being set against science or inquiry that is aimed at truth, is a part of it.

Another way in which instinct is a part of science is as follows:

> [W]hen one fact puts a person in mind of another, but related, fact, and on considering the two together, he says to himself 'Hah! Then this third is a fact', . . . it is by *instinct* that he draws the inference. (MS 682, p. 19, 1913)

If you feel that an inference is correct, that feeling is an 'instinct'. You have no overt *reason* to accept the conclusion – it just comes upon you. Peirce also makes this kind of point by saying that instinct is aligned to our habits of reasoning. Reasoning 'is the principal of human intellectual instincts . . . reasoning power is related to human nature very much as the wonderful instincts of ants, wasps, etc. are related to their several natures' (MS 682, pp. 8–9). Our instinctive and habitual cognitive skills guide our inquiries. Of course, these habits can be flawed, but we nonetheless rely on them until they are shown to be flawed – until we have evidence that they lead us astray or until we can explain what is wrong with them. If we are to continue to inquire, we must assume that our stock of habitual cognitive skills is reliable. Peirce is crystal clear that something's being such a regulative assumption of inquiry does not mean that it is true. But something's being a regulative assumption of inquiry does entail that we should believe it and that we should construct our philosophy so as to make room for its truth.

Finally, and most importantly, instinct is, for Peirce, also aligned with that which is not doubted – that which forms our 'commonsense' corpus of background beliefs. Writing in an entirely general way about belief and inquiry, Peirce says that 'the pragmatist will accept wholesale the entire body of genuine instinctive beliefs without any shade of doubt, tossing aside the toy doubts of the metaphysician as unworthy of a mature mind'.[12] This is the very point Peirce and Levi share regarding our background corpus of belief: It is Levi's epistemological infallibilism.

Here is how Peirce ties instinct to our background body of belief. He argues that even if you think, generally, that trusting instinct is 'treacherous and deceptive', if you don't doubt something and have never doubted it, you will believe it. Thus, 'that which instinct absolutely requires him to believe, he must and will believe it with his whole heart'. If something seems perfectly evident, you can try as you will to criticise it, but you will be obliged to believe it. 'Commonsense' and 'instinct', for Peirce, are interchangeable: They refer to 'those ideas and beliefs that a man's situation absolutely forces upon him' (CP 1.129, 1905). They are what the whole of past experience has put into place. They form our background body of belief.

These thoughts of Peirce's hold not just for science but also for ethics. In ethics, we must go on instinct or on the status quo. For what the whole of past experience has put in place is what is least likely to lead us astray when we are pressed to make a quick judgement. Ethics and science adopt the same method – they both require that our background beliefs are taken to be true and to be the standard for serious possibility. It is just that the ethical decisions often need to be made in a compressed time frame in which the gradual revision of those background beliefs is not an option.

Lest the reader not be convinced, it is helpful to look at the places in which Peirce's apparent gap between science and vital matters is most stark. One such place is the Cambridge Lectures of 1898, where we find this thought:

> We do not say that sentiment is *never* to be influenced by reason, nor that under no circumstances would we advocate radical reforms. We only say that the man who would allow his religious life to be wounded by any sudden acceptance of a philosophy of religion or who would precipitately change his code of morals, at the dictate of a philosophy of ethics, – who would, let us say, hastily practice incest, – is a man whom we should consider *unwise*. The regnant system of sexual rules is an instinctive or Sentimental induction summarizing the experience of all our race. (CP 6.633/RLT: 111)

[12] MS 329, p. 12, 1904; see also CP 5.445, 1905.

Parker[13] notes that we also find these thoughts in James's work. In 'The Moral Philosopher and the Moral Life', James extends Peirce's view of truth to ethics, arguing that society may be seen as a long-running experiment aimed at identifying the best kind of conduct. Its conventions thus deserve respect. Our background beliefs capture the experience of generations. James thinks that 'ethical science is just like physical science, and instead of being deducible all at once from abstract principles, must simply bide its time, and be ready to revise its conclusions from day to day' (p. 208). Peirce too is clear that, while the ethical deliberator might often be hesitant to revise her beliefs quickly, such hesitancy is not always justified:

> Like any other field, more than any other [morality] needs improvement, advance.... But morality, doctrinaire conservatist that it is, destroys its own vitality by resisting change, and positively insisting, This is eternally right: That is eternally wrong. (CP 2.198, 1902)

For both Peirce and James, ethical judgements are connected to experience in the way that all of our genuine judgements are: '[J]ust as reasoning springs from experience, so the development of sentiment arises from the soul's Inward and Outward Experiences' (CP 1.648, 1898). As with every other kind of experience, '[t]hat it is abstractly and absolutely infallible we do not pretend; but that it is practically infallible for the individual – which is the only clear sense the word "infallibility" will bear ... *that* we do maintain' (CP 1.633, 1898). This is exactly Levi's usage of 'infallibility', I contend.

We have seen that, for Peirce and for Levi, in any domain of serious inquiry we take our body of background belief to be practically infallible, until we need to revise a commitment. As Peirce puts it, our 'instinctual' and common-sense beliefs are subject to revision, but they are held firm until experience prompts that revision (CP 5.444, 1905). That is the Peirce I want to focus on. Ethics and science are in the same boat, relying on deeply held, but revisable, background beliefs and habits. Instinct has a positive and essential role to play in science and in ethics.

Enough has been said about how reliance on instinct does not distinguish ethical matters from other matters. Now we need to focus on the other aspect of Peirce's problematic point: the contentious view of science, in which belief is out of place. In the Cambridge Lectures we find Peirce saying that while the scientist can wait five centuries for an answer and thus does not believe his theories, the deliberator in ethics needs an immediate answer and thus has *beliefs* or commitments on which to act. But in these lectures Peirce see-saws

[13] Parker 1998, p. 50.

between the idea that belief has no place in science and the idea that it does.[14] First, he says:

> I would not allow to sentiment or instinct any weight whatsoever in theoretical matters, not the slightest. . . . True, we are driven oftentimes in science to try the suggestions of instinct; but we only *try* them, we compare them with experience, we hold ourselves ready to throw them overboard at a moment's notice from experience. (CP 1.634)

This is the 'no belief in science' side of the see-saw. We are not ready to act on belief in science. Science

> merely writes in the list of premisses it proposes to use. Nothing is *vital* for science; nothing can be. Its accepted propositions, therefore, are but opinions at most; and the whole list is provisional. The scientific man is not in the least wedded to his conclusions. He risks nothing upon them. He stands ready to abandon one or all as soon as experience opposes them. (CP 1.635)

But in the next breath, Peirce says that some of the scientist's conclusions are called 'established truths', 'propositions to which no competent man today demurs' (CP 1.635). Established truths are the background beliefs that we take for granted, the beliefs against which we judge new hypotheses. They are what the pragmatist focusses on. Peirce does indeed think that belief has a place in science.

Another source of the contentious view of science is the 1898 'Detached Ideas on Vitally Important Topics'. It is here that Peirce offers us his famous metaphor that science is walking upon a bog. The reason it can say only 'this ground seems to hold for the present. Here I will stay till it begins to give way' is that science always starts with an abductive inference. For Peirce, the conclusion of an abductive inference isn't to be believed; it is a mere conjecture. But

> [a]fter a while, as Science progresses, it comes upon upon more solid ground. It is now entitled to reflect: this ground has held a long time without showing signs of yielding. I may hope that it will continue to hold for a great while longer. (CP 5.589)

[14] These lectures are not the best place for discerning Peirce's considered view about science and vital matters. He was extremely irritated at James, who had charitably set up the lectures so that Peirce might quite literally be able to put a bit of food on his plate. On learning that Peirce intended to address technical questions of logic, James asked him to 'be a good boy and think a more popular plan out'. Perhaps he could rather speak about 'separate topics of a vitally important character'. Peirce, struggling no doubt with the shame of having to be rescued by James and having been shut out of an academic job by Harvard, pours scorn on the Harvard philosophers for their lack of training in logic and says with heavy sarcasm that he will indeed restrict himself to 'vital matters' – an area untouched by reasoning.

We can now use the hypothesis or conjecture in practice – we can believe it and we can act on it. For it no longer rests on a mere abduction. It has been inductively supported. Peirce says, 'In other words there is now reason to believe in the theory, for belief is willingness to risk a great deal upon a proposition'. Despite the fact that the scientist knows that experience and argument may at some point upset such beliefs, 'we call them in science established truths, that is, they are propositions into which the economy of endeavor prescribes that, for the time being, further inquiry shall cease' (CP 5.589).

Another source of the contentious view is Peirce's 1902 application to the Carnegie Institute, pleading for funds so that he could write his grand work on logic. There are many drafts of this application in the Peirce Papers, and they show very clearly that Peirce did not have a settled view about the matter in question. Perhaps his doubts are best expressed on page 54 of some of the drafts, where we have him saying that the scientist is in a bind, a 'double position':

> As a unit of the scientific world, with which he in some measure identifies himself, he can wait five centuries, if need be, before he decides upon the acceptability of a certain hypothesis. But as engaged in the investigation which it is his duty diligently to pursue, he must be ready the next morning to go on that hypothesis or reject it . . . he ought to be in a double state of mind about the hypothesis, at once ardent in his belief that so it must be, and yet not committing himself further than to do his best to try the experiment.[15]

What a wonderful statement of the problem. The inquirer (*any* inquirer) must be ready to believe and to act on the belief, knowing full well that it might not be true. Belief is not *out of place* in science; it is just tempered by knowing that revision might in the future be necessary. The scientist must believe, but be constantly aware that in subsequent steps, her belief might be overturned by the surprise of experience. In both scientific and moral matters, we have cherished beliefs that are nonetheless responsive to or sensitive to experience. In ethics, as in science, we act on our experience-driven background beliefs, while realising that they might yet be overthrown by further experience.[16]

This is a perfect statement of the tricky path on which the pragmatist must tread. And it is a perfect statement of Peirce's relation to Levi. While both share the anti-Cartesian view that we fully believe what we do not doubt, Levi emphasizes the fullness of that belief. Then he struggles with how to cope

[15] MS L75, pp. 53–5, of the first 88-page variant.

[16] See Misak 2000 for a discussion of how experience plays a role in ethics.

with the fact that agents often exchange that full commitment for something incompatible with it. Peirce, on the other hand, focusses on the kinds of things that can upset or unfix full belief – the surprise of experience, in its many forms. That's why Levi's title *The Fixation of Belief and Its Undoing* is a better title for Peirce than his own 'The Fixation of Belief'.

REFERENCES

James, William. 1979. "The Moral Philosopher and the Moral Life." In *The Will to Believe and Other Essays in Popular Philosophy*. Cambridge, Mass.: Harvard University Press. First published in 1897.

Levi, Isaac. 1980. *The Enterprise of Knowledge*. Cambridge, Mass.: MIT Press.

Levi, Isaac. 1984. "Messianic vs. Myopic Realism." In P. D. Asquith and P. Kitcher (eds.), *PSA: Proceedings of the 1984 Biennial Meeting of the Philosophy of Science Association*, pp. 617–36. East Lansing, Mich.: Philosophy of Science Association.

Levi, Isaac. 1991. *The Fixation of Belief and Its Undoing: Changing Beliefs through Inquiry*. Cambridge: Cambridge University Press.

Levi, Isaac. 2004. *Mild Contraction*. Oxford: Oxford University Press.

Misak, Cheryl. 1987. "Peirce, Levi, and the Aims of Inquiry." *Philosophy of Science* 54, no. 2: 256–65.

Misak, Cheryl. 1991. *Truth and the End of Inquiry: A Peircean Account of Truth*. Oxford: Clarendon.

Misak, Cheryl. 2000. *Truth, Politics, Morality: Pragmatism and Deliberation*. London: Routledge.

Misak, Cheryl. 2004. *Truth and the End of Inquiry: A Peircean Account of Truth*, 2nd expanded ed. Oxford: Clarendon.

Parker, Kelly. 1998. *The Continuity of Peirce's Thought*. Nashville: Vanderbilt University Press.

Peirce, Charles Sanders. 1958. *Collected Papers of Charles Sanders Peirce*, vols. i–iv edited by C. Hartshorne and P. Weiss (1931–35); vols. vii and viii edited by A. Burks. Cambridge, Mass.: Belknap.

Richardson, Henry. 1994. *Practical Reasoning about Final Ends*. Cambridge: Cambridge University Press.

2

Is Pragmatist Truth Irrelevant to Inquiry?

André Fuhrmann

> Such would be the scope of pragmatism – first, a method; and second, a genetic theory of what is meant be truth.
>
> *William James, Pragmatism (1907)*

1. INTRODUCTION

In 1909 William James published a sequel, *The Meaning of Truth*, to his famous lectures on pragmatism of 1907. This new book opens with the sentence:

> The pivotal part of my book named *Pragmatism* is its account of the relation called "truth" which may obtain between an idea (opinion, belief, statement, or what not) and its object. (p. v)

Judging by the fact that he continued to focus on the concept of truth, James can have felt no remorse at having made the theory of truth the pivot of his presentation of pragmatism.[1]

Apparently in strong disagreement with James, Isaac Levi writes that "pragmatism is seriously misrepresented as a philosophical outlook when attention is focused on the pragmatic theory of truth (whatever it may be)" (2002). Yet Levi agrees with the classical pragmatists, including James, that pragmatism is *first* a method and *second*, perhaps, a theory of truth. But Levi adds: If the second cannot be had without sacrificing the first, then we should let go of the idea of a distinctively pragmatist theory of truth. And he further adds: The second cannot be had without sacrificing the first. More specifically, Levi argues that the notion of truth that Peirce and James tried to describe as the ideal limit of converging inquiries does not square with the role that truth plays in pragmatically conceived inquiry. Truth in the limit

[1] **pivot** *n*. a pin on which anything turns: . . . that on which anything depends or turns: . . . . – *adj*. **piv′otal** . . . (*Chambers Dictionary*, 1983).

is, according to Levi, either a historical appendix to pragmatism, not nurtured by distinctively pragmatist concerns, or, if it is to be plugged into the theory of inquiry, a corrosive agent within the doctrine. In either case, modern pragmatists better live without it.

Levi's critical view of truth in the limit and its place in inquiry will be the topic of this chapter. In section 2 I identify a systematic equivocation in James's usage of 'truth'. Apart from a notion of truth that is internal to a given inquiry, James needs a further, absolute notion of truth that transcends particular inquiries. But such a notion invites the suspicion that it will be sieved out by the pragmatist criterion for a "workable" concept. The suspicion is deepened, in section 3, by a forceful argument that Levi has advanced and that serves as a cornerstone for his theory of inquiry. In brief, the argument states that if absolute truth is the goal of inquiry, beliefs once acquired should never be given up since, after all, they *might* be (absolutely) true. A Jamesian response to that argument will crucially depend on two hypotheses: First, under suitable conditions absolute truth will remain accessible, even if part of it may temporarily be lost; and second, under suitable conditions distinct lines of inquiry can be pushed to a unique limit. The conditions that govern the first hypotheses are investigated in section 4; those mentioned in the second hypothesis are the subject of section 6.

In section 4 I argue that an inquirer can be confident that absolute truth will eventually be within reach only to the extent that all *possible* evidence is made available to him. This condition is, of course, actually not fully realizable, but it provides the inquirer with a reason to actively search evidence concerning even those hypotheses that he or she currently considers to be settled. Still, if we define truth as the ideal limit of inquiry under the condition of unrestricted access to evidence, there remains the question as to the uniqueness of such a limit.

Section 5 is an interlude on the role of two notions of possibility in ongoing inquiry. Levi has argued that there can be no relevant role in inquiry for a notion of possibility other than the one that is naturally provided by a corpus of beliefs. Levi calls this notion of possibility, internal to a given corpus of beliefs, *serious possibility*. What the agent believes to be false is, from his current point of view, not a serious possibility. It might be a possibility in some other sense, but such a notion of possibility plays no role in justifying changes of beliefs at the next step of inquiry. I shall argue, as a corollary to section 4, that "metaphysical" possibility must play such a role, if inquiry aims at absolute truth.

The question of uniqueness is examined in section 6. James is keenly aware of a necessary characteristic of inquirers that seems to mar all hope for convergence. He calls this aspect “temperament”: a set of theoretical preferences that command how evidence is to be processed. For a viable notion of absolute pragmatic truth, it is thus not enough that inquirers have access to the same complete corpus of possible evidence; they also need to process evidence in the same way. Convergence of inquiries requires convergence of temperaments. By identifying a representation of temperament in the theory of belief change, I sketch an idea as to what convergence of temperament might mean. The role of absolute truth in inquiry is exactly this: It provides a reason for changing one’s theoretical preferences. Without concern for convergence it would never be rational to change our theoretical preferences. This is the conclusion drawn in section 7.

2. DAILY AND ABSOLUTE TRUTH, ACCORDING TO JAMES

When James applied the pragmatic maxim to the notion of truth, he asked in characteristic fashion: What “practical difference” does it make to pronounce a belief (opinion, idea, statement, etc.) true? James’s answer was that true beliefs “work.” The answer infuriated Russell and others – and was probably calculated to do so. It immediately invited cheap criticisms of caricature versions of pragmatism. Such criticisms could easily be dispelled by a more careful characterization of a “working belief.” But at least one type of criticism struck a deeper chord. Beliefs do their work – or don’t – only within a given context. If the context changes, the belief may work very differently. No amount of fine-tuning the notion of “working” or “expediency” – to use another of James’s favorite terms – can overcome this fundamental relativity.

But truth is not relative in this way. Whether a belief is true or not cannot depend on the circumstances in which it is held: which other beliefs are currently held, which kind of cognitive stress is suffered at the time, how much evidence is currently available, which kinds of theoretical preferences prevail, and so on. Truth is absolute. This was at least James’s view, when pressed. (It was not Dewey’s view, when pressed.)

But how can truth be absolute while at the same time being characterized in terms that make it highly sensitive to context? James sought a way out by introducing a systematic equivocation: There are *daily truths* and there is *Truth at the End of Days*. Daily truths are what we arrive at in our daily practice of inquiry; the final Truth is what awaits us when all inquiry comes to an end. Thus daily truths are temporary and relative, while Truth is final and absolute. More specifically, for James the absoluteness of Truth consists in its *stability* under the impact of further experience and in its eventual *invariance* to differences of points of view. Although stability and invariance are two distinct elements in the definition of truth, James frequently runs the two together in the idea of *convergence*:

> The "absolutely" true, meaning what no farther evidence will ever alter, is that ideal vanishing-point towards which we imagine that all our temporary truths will some day converge. (1907, p. 222f.)

The idea of convergence is also expressed in another passage where he considers the question of in which sense he could claim his theory of truth to be true. (Apparently, some critics believed that he could not.) He writes:

> I expect . . . that the more fully men discuss and test my account, the more they will agree that it *fits*, and the less will they desire a change. I may of course be premature in this confidence, and the glory of being truth final and absolute may fall upon some later revision and correction of my scheme. . . . To admit, as we pragmatists do, that we are liable to correction (even tho we may not expect it) *involves* the use on our part of an ideal standard. (1909, p. 142)

Absolute truth then plays an important role in James's pragmatism: It accommodates the demands for the nonrelativity of truth; it explains in which sense pragmatism itself can be said to be true; and it is essential for explaining the truth-aptness of beliefs about events that have left no available evidence.[2] James's introducing the notion of absolute truth thus closes the wound left by the charge of relativism.

Still, it cannot be denied that absolute truth receives a stepfatherly treatment in James's work. Absolute truth appears much less frequently on stage than daily truth, and its few appearances are hurried so as to give way to daily truth again. Absolute truth, James writes,

> runs on all fours with the perfectly wise man, and with the absolutely complete experience; and, if these ideals are ever realized, they will all be realized together. Meanwhile we have to live to-day by what truth we can get to-day and be ready to-morrow to call it falsehood. (1907, p. 223)

This passage plants the seeds of a disturbing suspicion. James's somewhat mocking remarks let appear absolute truth as a notion so exalted as to be irrelevant to daily problem solving. How then can absolute truth pass through the pragmatist's sieve for proper philosophizing?

[2] This problem is taken up at length in James 1909, particularly in the last chapter, "A Dialogue."

3. REOPENING THE WOUND: IS TRUTH IRRELEVANT TO INQUIRY?

In his lecture, “What Pragmatism Means,” James cites with approval Peirce’s formulation of what he calls “the principle of pragmatism”:

> Mr. Peirce, after pointing out that our beliefs are really rules for action, said that, to develop a thought’s meaning, we need only to determine what conduct it is fitting to produce: that conduct is for us its sole significance. And the tangible fact at the root of all our thought-distinctions, however subtle, is that there is no one of them so fine as to consist in anything but a possible difference in practice. (1907, p. 46)

The principle serves as a criterion of significance. A notion that makes no conceivable difference in conduct lacks significance. So what is, in pragmatic terms, the significance of “absolute truth”? If absolute truth has no pragmatically acceptable role in guiding conduct, then its taking part in a pragmatist theory of truth must be entirely gratuitous. Without any other kind of practical function, it remains a mere ad hoc device to fend off the charge of relativism. Its utility can at best be “on all fours” with that of “tender-mindedness’s hypothesis of an eternal perfect edition of the universe coexisting with our finite experience” (1907, p. 273). Yet, as James knew, philosophical dispute is tough, whence it is better not to count on the tender-mindedness of one’s opponents – or to leave such matters to the last chapter, as James (1907) did.

Even for James, who wrote so much about truth, pragmatism begins not with a view about truth but with a view about how to adopt beliefs. Already, the sequence of lectures on pragmatism (1907) shows that. James’s views about daily truth – and later absolute truth – are a corollary to the pragmatist theory of epistemic justification. As with most corollaries, it reports something worthwhile to say, but it says less than the distinctively pragmatist theorem from which it was derived. The content of the theorem cannot be recovered from that of the corollary. If it were not for the magic of words such as “utility,” “expediency,” and “cash-value,” what James has to say about truth can pass muster as a plain verificationist theory of truth. That pragmatism is just a kind of verificationism adorned with the magic of certain words is a comment that sometimes has indeed been made[3] but that is grossly besides the point. As Levi writes,

> pragmatism is seriously misrepresented as a philosophical outlook when attention is focused on the pragmatic theory of truth (whatever it may be). (2002)

[3] Perhaps most famously and influentially by Russell.

Instead, attention should focus on the pragmatist theory of epistemic justification. What is distinctively pragmatist in such theories is the rejection of the Cartesian doubt-belief model in favor of the Peircean belief-doubt model of inquiry. A characteristic consequence of adopting the belief-doubt model is a thesis about the objects of justification: not beliefs as such but changes of belief. Like no other pragmatist philosopher, Isaac Levi has identified this thesis as the backbone of philosophizing in the spirit of Peirce, James, and Dewey, and like no other philosopher before him, he has explored its consequences.

One of these consequences is that the suspicion mentioned at the end of the last section receives confirmation. What is new and distinctive in pragmatism, its theory of inquiry, does not cohere well with the idea of convergence to an ultimate truth, so says Levi. His argument is simple.[4]

Consider an agent *X* with a corpus of full beliefs *K*. Suppose that the way *X* changes from *K* to another corpus is dictated by his wish to get to *T*, the ideal corpus to be reached at the End of Days. Levi's argument is designed to show that *X*'s wish to get to *T* can have no command over how he changes his beliefs on the way from *K*. If it did, *X* would never contract, if fully rational. Thus belief systems could only grow, never be revised.

Let *p* be any proposition believed according to *K*. If *p* is a truth of logic or some other belief immune to revision, *X* should of course not contemplate giving up *p*. So suppose *p* is contingent. *X* knows two things: First, *p* might be part of *T*, and, second, if he now gives up *p*, he may never regain it. It follows that giving up *p* amounts to running the risk of getting further away from *T* without a subsequent reapproximation. So *X* should not contract. If *X* is fully rational, he won't.

The above argument is based on *X*'s concern with not missing out on truths according to *T*. This is the completeness part of *X*'s approximation to *T*. But were *X* concerned with truth at the End of Days, he would also aim at believing consistently with *T*. The above argument may be extended to show that under this supposition, contracting by *p* not only risks losing a truth but also risks importing an error.[5] For, once *p* is given up, $\neg p$ turns into a serious

[4] See Levi 1980, sec. 3.8; 1983, p. 126; 1991, sec. 4.10; 2002.

[5] According to Levi, what we do not believe we do not believe to be true. In seeking truth we would therefore either seek to believe what we already believe or we would seek to believe untruths – the first would make little sense, the second would be plain wrong. Thus, according to Levi, seeking truth cannot be an aim of inquiry. What we do aim at, however, is avoiding error. Error can be imported into our beliefs by involuntary belief formation – routine expansions – and needs correcting by subsequent contractions. Given this view of inquiry, the purpose of contraction would be defied by any supposition that implies that

possibility. In other words, if X now gives up p, he thereby opens himself to the risk of later expanding by $\neg p$. Not only would he thus miss out on a truth, but he would even trade in truth for error: He would go even further astray from T, a risk that X can and should avoid by not contracting.

The nucleus of the argument can be summarized in three lines:

(1) For any consistent p, p is possibly True.
(2) If p is possibly True, then p (once acquired) should not be given up. Hence,
(3) For any consistent p, p should not be given up.

As the conclusion (3) flies in the face of the plausible thesis that beliefs should be corrigible and since (1) is a truism, it is the premise (2) that must yield. But (2), so it seems, expresses the only way in which concern with Truth can play a role in ongoing inquiry. Thus, either in ongoing inquiry we should not be concerned with Truth or it needs to be shown that concern with Truth is not implemented by (2) but in some other, more acceptable way. So how does concern with Truth entail (2)?

A simple way to make (2) plausible is as follows. Let us think of the space of all possible belief sets, or states of inquiry, one of them, K^*. We may say now that a state K is *accessible* from a state H just in case there exists a sequence of expansions and contractions that lead from H to K. Many of these states of belief are maximally consistent: They tell a complete and consistent story of the world. One of them will tell that story Truly. Let that state of inquiry at the End of Days be T.

Supposing $p \in K^*$ (for some consistent p), X knows that p is part of some maximally consistent state but he does not know whether p is part of T. So it is possible, for all that X knows, that $p \in T$. So, moving from K^* to $K^* - p$, X (a) may or (b) may not distance himself from T. Since X does not know which of the two possibilities is actual, he needs to play safe. This is to say, if X's decision to contract or not to contract is dictated by a concern to aim at T, each of the scenarios (a) and (b) must be at least consistent with that aim. Possibility (b) clearly is consistent with the aim. So it remains to investigate what happens if X contracts in scenario (a). In that case $K^* - p$ will be further away from T than K^*. So, if X's concern to arrive at T requires

contractions may import error. It is this implication, then, which for Levi plays the of role of a premise, that reduces to absurdity the supposition that at each step of inquiry we are concerned with approximating to the Truth at the End of Days. For Levi, the initial step of the argument may score only as an *ad hominem*; it is the extension that tells, in his view, decisively against the mentioned supposition.

that *each step of inquiry be a step closer to T*, *X* should not contract. Thus we have derived premise (2) and so the above argument stands.

But the concern with aiming at *T* may be interpreted as allowing for a zig-zag trajectory from K^* to *T*. This is to say, contraction under the threat of (b) can be consistent with aiming at *T*, provided that *T* is still accessible from $K^* - p$. This interpretation faces two challenges. First, not all roads may lead to Rome. What assurance can there be that once a wrong turn is taken, *T* still remains accessible? This challenge will be taken up in the next section. Second, there may be many Romes. We have so far talked of truth as the result of pushing inquiry to *the* limit. But the definite article may be unwarranted. May inquiry not be pushed to more than one limit? This question is the subject of section 6 below.

4. GATHERING POSSIBLE EVIDENCE

We need a *guarantee* that *T* be accessible from $K^* - p$. For, if there were no such guarantee, contracting to $K^* - p$ would not be a case of playing safe: *X* might lose without having certainty that his loss will ever be recovered. That guarantee can be had quite trivially – at least on one interpretation of 'accessible'. For the operations of contraction and expansion make for a complete set of change operations in the sense that they jointly allow the inquirer to proceed from any given belief state to any other state. The safest method – wasteful in practice, though – is to contract to the empty state (or Levi's *urcorpus*) and to expand from there to the target state.

The suggestion just made will meet with an obvious objection. The concern that *X* may never get from $K^* - p$ to *T* may not stem from the worry that however he may change his beliefs, he will never recover *p*. That worry is ill-founded, as it has just been pointed out. Rather, *X* may never recover *p* by way of *justified* changes of beliefs. It may be that he will, as a matter of fact, never encounter evidence such that on the basis of that evidence *X* will be justified in changing his then present corpus *K* to $K + p$.

This is, in different guise, a problem familiar to pragmatists and to advocates of epistemic theories of truth in general. The world may just not offer sufficient evidence for every fact to become eventually known. James discusses the problem in various passages of his work, most extensively in the last chapter of *The Meaning of Truth* (1909), entitled "A Dialogue."[6] The dispute is about propositions concerning past events of which no evidential

[6] For a more detailed discussion of James's position and a comparison with Wright's (1992, 1998, 2001) theory of superassertibility, see Fuhrmann 2004a, forthcoming.

traces exist and will ever exist. James's interlocutor mentions as examples facts of "antediluvian planetary history" (154). Let p stand for any fact of this kind. The antipragmatist of the dialogue poses the following dilemma. On the one hand, it is admitted that p has a definite truth value. But then that truth value cannot be settled by appeal to justification now or in the future. For insofar as justification needs evidence, there can be none by our hypothesis. Thus, if p is true, then its truth cannot consist in what we have now or in the future have reason to believe – there is and will be no such reason. Alternatively, on the other hand, it may be said that under the hypothesis, p can be neither true nor false; propositions about events that leave no evidential traces are not truth-apt. This horn of the dilemma James immediately rejects, and so the ensuing discussion turns to the first horn.

Here is James's response:

> The truth of an event, past, present, or future, is for me only another name for the fact that *if* the event ever *does* get known, the nature of the knowledge is already to some degree predetermined. The truth which precedes actual knowledge of a fact means only what any possible knower of the fact will eventually find himself necessitated to believe about it. . . . This seems to me all that you can clearly mean when you say that truth pre-exists to knowledge. It is knowledge anticipated, knowledge in the form of possibility merely. (1909, pp. 157f.)

Given that truth and knowledge are wedded in the Jamesian manner, then

> wherever knowledge is conceivable truth is conceivable, wherever knowledge is possible truth is possible, wherever knowledge is actual truth is actual. Therefore when you point your first horn at me, I think of truth *actual*, and say it doesn't exist. It doesn't; for by hypothesis there is no knower, no idea, no workings. I agree, however, that truth *possible* or *virtual* might exist, for a knower might possibly be brought to birth; and truth *conceivable* certainly exists, for, abstractly taken, there is nothing in the nature of antediluvian events that should make the application of knowledge to them inconceivable. (1909, p. 158)

What James insists on in this passage is that we identify in the right way the range of inquiries that are claimed to converge on a given truth-claim. As inquiries are triggered by evidence, lack of it naturally impedes progress or may even block inquiry entirely from the beginning. But even if an inquirer fails to take note of certain evidence that is actually available to him, he or someone else actually could. In that case inquiries that take full account of all actually available evidence virtually exist. Any stable verdict of such inquiry is what James calls a *virtual* (or possible) *truth*.

But taking account of all actually available evidence may not be enough. We can imagine "mute" events, such as the aforementioned antediluvian events, whose evidential traces have vanished before inquiry could have begun or that may have never left any evidence. Another example of the latter kind would be a visit of superroaches, that is, cockroaches that devour all evidence before leaving. In such a case there *is* no evidence that an inquirer actually can take note of, however hard he may try. But there *could* be such evidence. We can always imagine a situation in which an event of this kind has occurred but ceases to be mute. Thus the superroaches may have visited, but their usual cleaning-up routine may have failed. Or time might not have wiped out all traces of certain, actually mute events that occurred in some prehistorical period. So inquiries that make use of such counterfactually available evidence are conceivable; their stable verdicts make up the class of *conceivable truths*.

Here and elsewhere James relies on the assumption that no event can be intrinsically mute; that muteness can happen to an event only because of circumstances. For James such circumstances are roadblocks to inquiry that can, as a matter of possibility, be cleared out of the way. The assumption is difficult to assess.[7] To be sure, from a pragmatist point of view, the notion of an intrinsically mute event, an event that can have no possible impact on inquiry, is an evident candidate for getting caught in the sieve of the pragmatic sense criterion. This, however, is an argument that is likely to work only on home turf; it would be better if an argument against intrinsically mute events could be produced that did not rely on such a strong, specifically pragmatist principle. Let us for now grant that assumption. It remains to be seen where James's account of absolute truth as the outcome of *conceivable* inquiry leaves us with the question as to whether justifiable progress to absolute truth is consistent with sometimes contracting one's corpus of beliefs.

We have seen that if we define accessibility between belief states in terms of *legitimate* belief changes, the condition that the absolute truth, T, be in this sense accessible from any state of belief ceases to be guaranteed. The guarantee fails because in the actual course of events there may be evidence missing to effect the necessary expansions in a *legitimate* way. James's response to the problem of mute events suggests that we balance the restrictive move from

[7] Fitch's (1963) argument may be interpreted as showing that certain epistemic events are necessarily mute. A Jamesian therefore needs to respond to Fitch's paradox. Replies in the style of Edgington (1985) appear to be very much in the spirit of the passage from James cited above.

logically possible to legitimate belief changes by a liberating move from factually to possibly available evidence.

This balancing act does indeed serve the purpose at hand. For, suppose that event *e* has occurred and let *p* be the belief that *e* has occurred; so $p \in T$. Assuming that *e* cannot necessarily be mute, there exists *possible* evidence that *can* be encountered at any given point *K* in inquiry such that given such evidence the inquirer would be justified in expanding to $K + p$. Having removed *p* from his corpus prior to reaching *K* does not impede *X*'s *possible* progress to *T*. But now, so it must seem, we are even further afar from the idea that progress to absolute truth can be an aim in *actual* inquiry. Answering the second challenge presented at the end of section 3 will give us some material at hand for finally suggesting a role for concern with ultimate truth in ongoing inquiry (section 7). But first some remarks on "metaphysical possibilities" and their role in inquiry are in order.

5. INTERLUDE ON SERIOUS AND METAPHYSICAL POSSIBILITY

We have considered contraction under the threat of unfavorable circumstances. Let *p* be an arbitrary candidate for contraction. Then *p* may be absolutely true, that is, *p* may be part of *T*, the ideal corpus of beliefs to be obtained at the end of inquiry. Contracting by *p* may thus not result in coming closer to *T*; on the contrary, it may put more distance between one's present beliefs and *T*. We have seen that this threat need not impede contracting, as long as reapproximation to *T* remains possible. Such reapproximation is indeed guaranteed, provided the agent *X* will subsequently be in a position to take into account all possible evidence concerning *p*. Confidence in reapproximation is then justified to the extent that the agent succeeds in satisfying this proviso. Thus, assuming that the agent seeks approximation to absolute Truth, he must strive for taking into account as much evidence concerning *p* as he possibly can. After having removed *p* from his corpus, the agent must remain open-minded as to *p* and must continue to remain so even after having opted for believing $\neg p$ at some later stage of inquiry. Without such open-mindedness, contraction may initiate a path of error without return.

According to Levi, a corpus *K* that contains full belief in $\neg p$ leaves no serious possibility for *p*. As Levi has repeatedly pointed out, this is perfectly compatible with *X*'s realizing that in some future state of inquiry he may come to believe *p* or, more generally, that *p* may be part of some state of belief other than *K*. The agent is aware of the fact that most of his beliefs are *corrigible*. In other words, the fact that *X* rules out *p* as a serious possibility does not preclude that he recognizes *p* as a candidate for belief. Let me use the term *metaphysical possibility* for a candidate for belief.[8] Besides Levi's notion of serious possibility, which is always relative to a given corpus, there is thus another notion of possibility that is not relative to a corpus[9] and that the agent can recognize.

Levi has denied that the appreciation of metaphysical possibilities – in stark contrast to that of serious possibilities – is of any relevance for conducting inquiry.

> [E]valuations of hypotheses with respect to serious possibility have an important role to play in deliberation and inquiry. Metaphysical possibility lacks such clear role. It appears eminently expendable. (1983, p. 190)

According to Levi, that beliefs are in general corrigible is one of many facts about beliefs which an agent will in general be aware of. But the awareness of this fact is, like that of many others, irrelevant when it comes to justifying the passing from one state of belief to another. It must be borne in mind that Levi is thinking here of inquiry as concerned with avoiding error only at the next step to be taken. Denying metaphysical possibility a relevant role in inquiry in this sense may well be compatible with assigning it such a role in inquiry in the sense of Peirce or James. To the latter I turn now. I argue that recognizing metaphysical possibilities in this sense must be a relevant factor in any inquiry that is concerned with approximation to the ultimate truth.

As pointed out above, recognizing approximation to T as the aim of inquiry requires that one continue searching for evidence concerning even those beliefs that have been settled at a certain stage K of inquiry. If p is believed according to K, then $\neg p$ is, judged by K, not seriously possible. If we consider only judgments of serious possibility relative to K, then the issue as to whether p must be regarded as closed. But when contracting with a view to truth in the long run, we need reassurance that the case as to whether p may be reopened. This reassurance is given only by the corrigibility of one's beliefs – the metaphysical possibility of adopting contrary beliefs. If the agent had no certainty that the future course of experience possibly leads to reexamining his position as to p, then he could not be justified in removing p now under the threat of thereby losing a Truth now. Recognizing this possibility works

[8] I am not trying to import some independently fixed notion of metaphysical possibility. In Levi's terminology the notion of possibility I am adverting to here would perhaps best be rendered as *logical consistency with the urcorpus*. The urcorpus is, according to Levi, a body of absolutely unassailable beliefs. Levi deliberately does not enter into considerations on the nature of their unassailability. Likewise, I have no particular view to offer as to what might or might not be a candidate for belief or disbelief.

[9] Or, at most, invariably relative to the urcorpus; see the previous footnote.

like the right to a return ticket back to a point that promises a safer departure to Truth. Only with this right in hand can a belief change by contraction be justified. Thus, recognizing metaphysical possibilities is necessary for justifying belief change by contraction – *if* approximation to the ultimate truth is the goal of inquiry.

6. TEMPERING TEMPERAMENT

In section 4 we encountered some assurance that from any given state of belief we shall be able to reach a limit of inquiry in which all questions will be settled – undefeatably by further evidence. We are now taking up the second challenge mentioned at the end of section 3. Why should all inquiries head for the same limit? Is there any such thing as a unique ultimate state of inquiry?

This is, of course, a familiar challenge to theories that affirm in one way or other that different inquiries will eventually converge in their verdicts, if only allowed sufficient time and access to evidence. What is perhaps less familiar is that it is precisely James's theory of inquiry which attends to such aspects of inquiry that seriously undermine any confidence in convergence – even under the ideal condition of complete access to evidence.

What seriously interferes with the ideal of convergence is something James called *temperament*. Different inquirers, when confronted with the same evidence, may have different dispositions to react to that evidence. How one might adjust one's beliefs when realizing a certain tension between them or when faced with new evidence may depend not only on one's present beliefs, but also on how deeply entrenched they happen to be. It is mainly this second aspect that James has in mind when he writes of temperament. Speaking of philosophical inquiry, he writes:

> [H]is temperament gives [the philosopher] a stronger bias than any of his more strictly objective premises. It loads the evidence for him one way or the other, making for a more sentimental or a more hard-headed view of the universe, just as this fact or that principle would. He *trusts* his temperament. (1907, p. 7)

What James observes here applies not only to philosophical – though perhaps very strikingly so – but to any kind of inquiry.

In a sense, James's observation is at the heart of the theory of belief change.[10] If belief change did not need temperament, it would be trivial.

[10] I am thinking here of the theory of belief change as can be found, e.g., in Levi 1980 or Gärdenfors 1988.

Let p be a hypothesis presented to an inquirer X in a state K. First, presentation of p, even if well supported by evidence, does not automatically dictate a change of K. X may decide to discard the evidence supporting p – a first manifestation of temperament. Second, once X decides to adopt p, the adjustment of K called for is not in general logically determined by the new belief p and the old beliefs in K. Let K be the logical closure of $\{q, q \rightarrow \neg p\}$. If X wishes to adopt p as a new belief, then, to put it simply, he will have to remove either q or $q \rightarrow \neg p$ from K. Logic alone cannot determine which of these beliefs should go. A choice has to be made. Making such a choice is a second manifestation of temperament. If it were not for the necessity of making choices, the theory of belief change would be but a straightforward corollary to the theory of logical closure operations. In the literature on belief change the choice situation is usually represented either by entrenchment relations between beliefs or by choice functions on families of subsets of belief states that are maximally consistent with a given hypothesis.[11]

There is yet a third manifestation of temperament. The additional structure on belief states that is necessary to resolve choice situations cannot be assumed to be fixed throughout a sequence of belief changes. In other words, temperament may change. While this complicates the theory of belief change considerably, for James it harbors hope that attaining ultimate truth is possible, as I explain in the final section.[12]

Given the fact of temperament and its indispensability for effecting changes of belief, it is easy to see that in general there can be no guarantee that the belief states of two agents X and Y are headed in the limit for a unique state T, no matter how much evidence is being made available to them. This remains true even if we assume that both X and Y start from the same belief state, be it the urcorpus or some other corpus of beliefs. If X and Y are tempered differently, they simply may not react to poised evidence in the same way. Believing in convergence seems indeed an inexplicable albeit touching act of faith, as Levi once put it (1991, p. 163). Ironically, James emphasizes exactly those aspects of inquiry strongly suggesting that convergence is not to be expected.

[11] Grove's modeling in terms of systems of spheres may be considered a variant of the first alternative. As one might expect, the two representations are intimately connected. Rott 2001 contains the most complete study of these interconnections in the larger context of the general theory of preference and choice.

[12] In the literature on belief change, the problem of facing changing choice functions is known as the problem of iterated belief change. Iterated belief change is not much studied because there is little scope for making non-trivial general observations about the phenomenon. The few studies there are about the phenomenon consider iterations under highly simplifying conditions.

The problem is not that inquiry may not come to an end but that it may come to many ends.[13] If so, then there are three options. First, we may identify truth with the result of joining all limit inquiries. If each of the limits is complete, then truth is inconsistent; if the limits need not be complete, it would still border on the miraculous how their join could escape inconsistency.[14] Second, we may identify truth with what the limits of inquiry agree on. Then, unless they agree on everything, truth will be incomplete. Perhaps James is really committed to truth being incomplete. Yet he denies such commitment. Third, we may go back on the idea that truth coincides with the result of all inquiries in the limit.

This third option cannot here be interpreted as recommending that truth be identified with the result of one, distinguished inquiry. Apart from the notorious difficulty of distinguishing in a noncircular manner the inquiry that is to furnish the standard of truth, the idea simply does not serve the purpose still at hand. For we are yet to find assurance that Truth, *T*, be accessible from *any* belief state – particularly from certain imperfect ones. Only if Truth is accessible from everywhere can contraction under the threat of removing a possibly true belief be justified.

There is a simple way of guaranteeing the eventual coincidence of different inquiries. As we said above, the threat of divergence stems from the possibility that *X* and *Y* are differently tempered: *X* and *Y* may resolve differently choice situations that arise in episodes of belief change. If *X* and *Y* could be brought to exercise the same choices – if, in other words, they would eventually adopt the same selection function or epistemic entrenchment relation to determine candidates for removal – then, given sufficient exposition of evidence, their states of belief will indeed converge in the limit. So as to achieve convergence

[13] In Fuhrmann forthcoming, I have argued that inquiry can be represented in such a way that the conditions for algebraic fixpoint theorems in the spirit of Knaster and Tarski apply. The end of inquiry may then be understood as the first fixpoint of a function that drives changes of belief in response to evidence. But there really should be no need for excursions into fixpoint theory. That evidence can, in principle, be exhausted and that belief states stabilize thereby until an absolutely stable point is reached seems eminently plausible. Challenging the idea that inquiry can come to an end – though perhaps not to a unique one – is an act of ill-motivated skepticism.

[14] Note that this option is not of the kind that can attract a dialetheist (dialetheism is the view that there are true contradictions). There is no reason to expect that inconsistency afflicts only ultimate beliefs about liar sentences, the Russell class, or flying arrows – beliefs that, with some plausibility, could be held to be irresolvably paradoxical and thus both truth and false; cf., e.g., Priest 1995. As inconsistency would also arise from processing differently evidence concerning very ordinary beliefs, any belief could turn into a candidate for paradox. This, I take it, is too much even for a dialetheist.

of beliefs we need convergence of dispositions to believe or, as James put it, of temperaments.

James indicates in various passages that he considers tempering temperament indeed to be a necessary condition for progress in inquiry. Temperament must not be allowed to influence the result of inquiry in uncontrollable ways; it must itself be brought within the scope of inquiry. Thus James writes about controlling temperament in philosophical inquiry:

> [I]n the forum [the philosopher] can make no claim, on the bare ground of his temperament, to superior discernment or authority. There arises thus a certain insincerity in our philosophic discussions: the potentest of all our premisses is never mentioned. [*1*.] I am sure it would contribute to clearness if in these lectures we should break this rule and mention it. . . .
>
> [*2*.] Most of us have, of course, no very definite intellectual temperament, we are a mixture of opposite ingredients, each one present very moderately. We hardly know our preferences in abstract matters; some of us are easily talked out of them, and end up by following the fashion or taking up with the beliefs of the most impressive philosopher in our neighbourhood, whoever he may be. [*3*.] But the one thing that has *counted* so far in philosophy is that a man should *see* things, see them straight in his own peculiar way, and be dissatisfied with any opposite way of seeing them. There is no reason to suppose that this strong temperamental vision is from now onward to count no longer in the history of man's beliefs. (1907, pp. 8f.)

In the order of appearance there are three theses clearly discernible in this passage. First, theoretical preferences need to be examined, just as beliefs do. Second, if they remain shielded from inquiry, their consequences or mutual inconsistency may go unnoticed. Third, once theoretical preferences are adopted as the result of inquiry, they should be exercised unhesitatingly – they are then infallible from the inquirer's point of view.

Tempering temperament is certainly a valuable maxim. But exercising that maxim to the degree required for convergence is to enter an infinite regress. For it is precisely James's theory of inquiry that draws attention to the fact that belief change without exercising choices is impossible. Thus, if we bring our dispositions to choose within the scope of inquiry so as to possibly change them, we shall again need to choose.

There is a further complication here. Beliefs about cups and saucers and beliefs concerning how one might change beliefs about cups and saucers are of very different kinds. They are so different that it is plausible to suppose that they cannot be brought within the same axis of inquiry. I am adverting here to a fact well known to students of belief change. If assertions about what one would believe after contracting a corpus K are systematically included in the very same corpus K, then contraction will be monotonic in the sense that (for any belief A) if $K \subseteq H$, then $K - A \subseteq H - A$. But if contraction is monotonic, and if it satisfies certain basic principles, then there can be no belief states that are consistent with three arbitrarily chosen pairwise inconsistent beliefs.[15] So, beliefs about how to change beliefs in a corpus K need to be quarantined in a metacorpus $M(K)$, and so on up the hierarchy. Changing temperament, then, as required for convergence, amounts to changing belief states in the hierarchy of metacorpora for a given belief state. Thus convergence to a single ultimate corpus T would require simultaneously "horizontal progress" about cups and saucers and "vertical regress" about how to change one's opinions about cups and saucers.

7. THE DEMANDS OF ABSOLUTE TRUTH ON ONGOING INQUIRY

How such simultaneous belief change along the horizontal and the vertical axis is to be achieved, and whether it can be achieved at all, no one has yet investigated in any sufficient detail. But, as argued above, approximation to a unique ultimate state of inquiry that can serve as an ideal standard of absolute truth requires exactly this. It requires in particular that theoretical preferences (the vertical axis) be subjected to change until all idiosyncracies are leveled out to a common and stable standard of how to change one's beliefs along the horizontal axis. This, at any rate, I believe was the view James was committed to.

It follows that ultimate truth can be achieved only inasmuch as dispositions to change one's beliefs become uniform across the relevant community of inquirers. Here we finally encounter the pragmatic role of absolute truth in inquiry. Aiming at absolute truth is not only consistent with contracting belief states along the horizontal axis. Absolute truth as a goal of inquiry also exercises a normative command on how each step of belief change is to be effected by bringing the option of changing theoretical preferences into view. If an inquirer X were not concerned with absolute truth, he would have no reason to change his theoretical preferences. Every inquirer would be free to converge on his own personal limit of inquiry. In a community of stubborn inquirers there is no hope for attaining absolute truth. Inasmuch as X is concerned with truth, he *must* aim at aligning his theoretical preferences with those of others. This, at any rate, seems to be the critical role of

[15] Beliefs A and B are *inconsistent*, if the fact that A entails $\neg B$ is a theorem of the logic that defines the notion of closure used in the condition that belief states be logically closed. The result reported in the text is due to Gärdenfors; see 1988, sec. 7.4.

the *community* of inquirers that the classical pragmatists were eager to bring into focus. Just as appeal to possible lines of inquiry is thought to eliminate the possibility of truth-value gaps due to circumstantial lack of evidence, so appeal to a community of possible investigators is meant to clean up truth-value gluts as a result of processing possible evidence according to different theoretical preferences. Whether this pragmatist strategy can succeed appears to be still an open question.

REFERENCES

Edgington, D. 1985. "The Paradox of Knowability." *Mind* 94: 557–68.

Fitch, F. 1963. "A Logical Analysis of Some Value Concepts." *Journal of Symbolic Logic* 28: 135–42.

Fuhrmann, A. 2004a. "Absolute pragmatische Wahrheit." In A. Fuhrmann and E. Olsson (eds.), *Pragmatismus heute*. Frankfurt a.M.: Ontos.

Fuhrmann, A. Forthcoming. "James Unpeirced and Unfitched." *Cognitio* 5.

Gärdenfors, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, Mass.: Bradford/MIT.

James, W. 1907. *Pragmatism*. New York: Longmans, Green.

James, W. 1909. *The Meaning of Truth*. New York: Longmans, Green.

Levi, I. 1980. *The Enterprise of Knowledge*. Cambridge, Mass.: MIT Press.

Levi, I. 1983. "Truth, Fallibility and the Growth of Knowledge." In R. S. Cohen and M. W. Wartovsky (eds.), *Language, Logic and Method*, 153–74. Dordrecht: Reidel. Page references to the reprint in Levi 1984.

Levi, I. 1984. *Decisions and Revisions*. Cambridge: Cambridge University Press.

Levi, I. 1991. *The Fixation of Belief and Its Undoing*. Cambridge: Cambridge University Press.

Levi, I. 2002. "Seeking Truth." In W. Hinzen and H. Rott (eds.), *Belief and Meaning: Essays at the Interface*, 119–138. Munich: Hänsel-Hohenhaus.

Priest, G. 1995. *Beyond the Limits of Thought*. Cambridge: Cambridge University Press.

Putnam, H. 1997. "James's Theory of Truth." In R. A. Putnam (ed.), *The Cambridge Companion to William James*, pp. 166–85. Cambridge: Cambridge University Press.

Rott, H. 2001. *Change, Choice and Inference: A Study of Belief Revision and Nonmonotonic Reasoning*. Oxford: Clarendon.

Wright, C. 1992. *Truth and Objectivity*. Cambridge, Mass.: Harvard University Press.

Wright, C. 1998. "Truth: A Traditional Debate Revisited." *Canadian Journal of Philosophy*, suppl. vol. 24: 31–74.

Wright, C. 2001. "Minimalism, Deflationism, Pragmatism, Pluralism." In M. Lynch (ed.), *The Nature of Truth*, pp. 751–87. Cambridge, Mass.: MIT Press.

3

The Knowledge Business

Philip Kitcher

I

Isaac Levi is heir to the great tradition of American pragmatism, and particularly to the strand of it that runs from the writings of Peirce on knowledge and method through the more scientifically oriented works of John Dewey. Levi's approach to epistemology and philosophy of science, begun in *Gambling with Truth* and further developed in *The Enterprise of Knowledge*, *The Fixation of Belief*, and more recent essays and monographs, offers a detailed and rigorous elaboration of a pragmatist picture. In what follows I am less concerned with the details than with the general picture, much of which I find both congenial and important.

Main lines of Levi's view descend from Peirce's belief-doubt model of inquiry. In particular, he continues the following themes:

1. The rejection of Cartesian thoughts to the effect that doubt is always appropriate; there have to be specific reasons to prompt doubt.
2. An emphasis on *improving* our stock of beliefs, rather than *grounding* them; the task for epistemology and philosophy of science is to identify ways in which inquiry should go better. (There are connections here to ideas of Otto Neurath and W. V. Quine.)
3. Functionalism about the standards that govern inquiry; the methods we develop should be well adapted to promote our goals; we do not arrive at these methods a priori, but discover them in the course of our investigations. (In the tradition of Dewey, Levi is also sympathetic to the thought that we might inquire into goals as well as into methods.)

It is a great pleasure to contribute to this volume in honor of Isaac Levi, teacher, colleague, and friend.

4. Recognition that we all adopt routines for adding new beliefs; the most obvious of these consist in our ways of expanding our corpus of beliefs through perception and in response to the information we receive from others.

These seem to me to be important insights.

In my judgment, epistemology ought to move beyond any search for foundations, worrying less about whether the place in which we have found ourselves is secure than about good ways for going on from where we are. Moreover, I see little point to the common philosophical game of pitting definitions of 'knowledge' against intuitions. Better, I suggest, to think about the goals at which inquiry aims and to adopt an explicitly functionalist perspective on methods.

I see no harm in explicating those goals, in part, by drawing on a notion of truth and taking the straightforward approach to the notion of error that views errors as false beliefs. So far, there is no conflict with pragmatism, ancient or modern. But the notion of truth I want to deploy is a modest correspondence theory, one that extends the familiar Tarski apparatus by conceiving of it as reducing the notion of truth to that of reference, and viewing reference as a natural relation between representations and parts of an independent nature. The classical pragmatists wanted to reject correspondence theories of truth because the versions of such theories prominent in their intellectual context were weighed down with a vast load of metaphysical assumptions. Our situation is different, and I believe that contemporary philosophy can articulate a modest correspondence theory of truth that will avoid both the tangles into which pragmatist accounts of epistemic goals are so often led and also the dubious doctrines that pragmatists distrust. To articulate this point would require a whole paper, and since I want to focus on different questions, I shall simply break with pragmatism on this point and formulate some of my claims in terms of the notion of truth (understood in the modest correspondence terms I espouse).[1]

The questions I intend to explore stem from an aspect of Levi's preferred versions of 4. Although he embeds the individual in a community of knowledge-seekers, and although he recognizes the importance of those routines of expansion in which we borrow from other people, Isaac Levi

[1] The defense of this approach is begun in my essays "Real Realism: The Galilean Strategy" and "On the Explanatory Power Of Correspondence Truth." Levi is skeptical of the kind of theory of truth indicated in these papers, but I don't think our disagreement will materially affect the discussion on which I embark here.

remains a rugged individualist. At the center of his epistemological work is the changing belief corpus of a single cognitive agent. Insofar as there is any reference to community knowledge or to public knowledge, the community is conceived as if it were a massive individual. My aim is to explore what happens when we take a more thoroughly social approach to a framework very much like Levi's (with the difference, already noted, that it permits talk of truth). I hope to show that some important, and completely neglected, epistemological problems emerge.

II

I have already referred dismissively to the very un-Levite philosophical game of bouncing proposed definitions of 'knowledge' against ever more bizarre and recondite cases. My lack of interest in these activities is based both on a sense of how fragile the allegedly "intuitive" responses are and on the obvious point that, at best, these exercises would reveal only our current usage of 'know' and cognate terms. Instead, we should ask the explicitly functional question of what notion of knowledge might be apt for the purposes of our inquiries.

In a discussion that is both rich and concise, Edward Craig has investigated the function served by our current practices of attributing knowledge to others. He offers the plausible suggestion that we want the notion of knowledge to mark out those sources of information we can usefully employ. To put the matter in Levi's terms, the concept of knowledge plays a role in characterizing preferred routines for expansion, in that we want to adopt into our corpus those items of information delivered by subjects we identify as knowers.[2]

The most evident way to understand this function is to start with a state in which individuals can acquire information by following perceptual routines and in which they engage in face-to-face interactions that may involve transmission of information from one to another. We suppose that it is genuinely advantageous to borrow from another because of the high cost (or perhaps impossibility) of investigating everything that matters for oneself, but only if the borrowing does not incorporate error (false belief) into the corpus. Hence the introduction of a notion of knowledge to demarcate the occasions on which

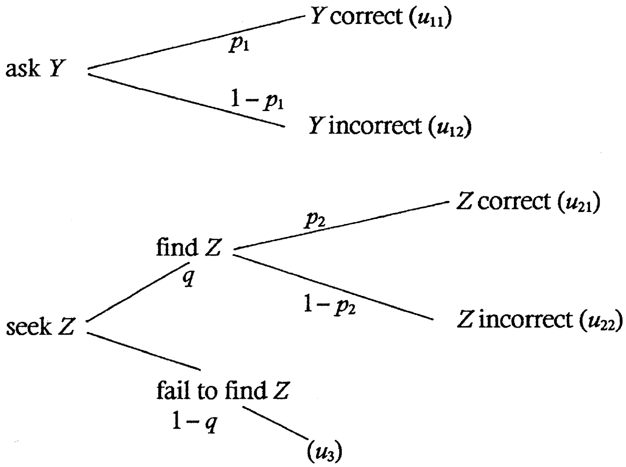
[2] Levi considers two modes of expansion, *routine* expansion and *deliberative* expansion. In what follows, I am concerned with routine expansion and with decisions about social routines.

borrowing is a good idea: X resolves to take on the information delivered by Y in just those cases in which X identifies Y as a knower (or as a knower with respect to the type of issue that is in question).

Plainly, it would serve the purposes that agents have in mind if those from whom they borrow were to believe truly. But this cannot be the *basis* for one person's selection of another as a source: The other would be redundant if the borrower could already assess the truth of the belief transmitted. Appropriate sources, knowers, must then be marked out in some other way, as individuals with a *tendency* to arrive at true beliefs (or at true beliefs about a range of matters); they must "track truth" or be reliable, in the sense of having a high probability of arriving at true beliefs (on the pertinent topics). Evidently, this functional approach accounts for the attractiveness of reliabilist epistemology.

Two obvious questions arise about how to make this approach more definite: One concerns the threshold to be set for reliability; the other involves the class of contexts (the reference class) across which we assess the agent's performance. Thinking about the functions of marking others as knowers helps with both. Standards of reliability for treating someone as a knower depend on the attributor's *purposes* and the attributor's *options*. For some purposes, provision of information may be so important that we are prepared to settle for a lower index of reliability; on other occasions, it is much more crucial to avoid error. (In noting this trade-off, clearly, I recapitulate a general point that Levi has emphasized from *Gambling with Truth* on.) Even here, however, the policy of marking others as knowers would best be articulated by looking for those available sources with the highest reliability index. Because we are interested in using the demarcation across contexts, it might still be valuable to recognize someone as a knower even though on some occasions there are available sources with higher reliability. The characterization of the routine for expansion, however, should be geared to the acquisition of information from the most reliable source available (assuming, of course, that both sources would provide information at the same level of specificity).

In addressing the question of specifying the threshold of reliability, I have introduced another hazy notion, that of availability. We can make this more precise by recognizing that the expansion of belief with respect to some issue about which one needs information can often be framed as a decision problem. Imagine a subject X who seeks the answer to a particular question. Another individual, Y, who might provide the answer, can be consulted without further ado. Alternatively, X might seek out a different source Z in hopes of increasing the chance of obtaining the correct answer. Assuming (for simplicity) that either potential informant, if consulted, will deliver some answer, X is in a predicament with the following structure:



As we would expect, even though Z is genuinely more reliable, that is, $p_2 > p_1$, the costs associated with seeking Z can depress the utilities so as to make consulting Y the preferred option. [Assume that $u_{21} = u_{11} - v$, $u_{22} = u_{12} - v$, $u_3 = p_1 u_{11} + (1 - p_1)u_{12} - v$; consulting Y is preferable if $p_1 u_{11} + (1 - p_1)u_{12} > q\, p_2(u_{11} - v) + q(1 - p_2)(u_{12} - v) + (1 - q)(p_1 u_{11} + (1 - p_1)u_{12} - v)$; which obtains if $v > q(p_2 - p_1)(u_{11} - u_{12})$.] This tells us that consulting that requires no work is preferable when the costs of search (v) are larger than the product of three terms: the probability of success in finding a more reliable source (q), the difference in reliability ($p_2 - p_1$), and the relative value of being right rather than wrong ($u_{11} - u_{12}$).

The specification of a threshold for reliability should be set by considering the options available to the subject. For X to treat Y as a knower in the simple scenario just outlined is for X to think that the costs of seeking out Z tell against preferring that option and that the expected value of consulting Y is positive ($p_1 u_{11} + (1 - p_1)u_{12} > 0$). To resolve the problem of delineating more precisely the class of contexts to be considered in assessing the reliability of others (the reference class problem), we should focus on the kinds of discriminations that the potential recipient of information can make. Imagine that X is concerned with whether or not to adopt the proposition expressed by Y's tokening of S, where S is an occasion sentence. Assume that X has beliefs about the relative frequency of Y's correct responses. If X can distinguish two types of context – those distinguished by the presence or absence of some factor F – in which Y's probability of correct tokening is taken to be different, then X would be incorrect in lumping the contexts together, unless X can assign no probability to the presence of F on the occasion of concern. If X knows that F is present on this occasion, then X's evaluation of Y's reliability should be the probability of correct tokening on occasions marked by F; if X assigns probability r to the presence of F on this occasion, then the assigned reliability of Y should be $rp_{\mathrm{F}} + (1 - r)p_{\mathrm{F}}$, where rp_{F} is the chance of responding correctly when F is present and $(1 - r)p_{\mathrm{F}}$ is the chance of responding correctly when F is absent. In effect, X should use the narrowest class for which probabilities can be assigned.

Analogously for eternal sentences. Here the task for X is to evaluate Y's rate of success within a class of propositions, and the task is best undertaken by picking the narrowest class to which probabilities can be assigned. If we are interested in gaining information about job growth during a particular period, we'll prefer to assess potential informants according to their rate of correct responses with respect to questions in labor economics rather than by their success in economics generally – although the latter would be a better measure than their success across the entire range of scientific statements.

Quite evidently, the reliability that the potential recipient of information assigns to a source might itself be the object of investigation. Sometimes it will be worth conducting a preliminary inquiry to arrive at a better version of the probability that the envisaged informant will give the correct answer, or even to come to recognize the sorts of distinctions that are best made in assessing rates of success. When the costs of improving one's assignment of reliability are too high, the preferred option for the subject will be to proceed with whatever assessments can be derived from the current belief corpus, even though the view from an Olympian perspective untroubled by considerations of cost might revise the judgment in consequential ways.

I believe, although I shall not develop the point here, that the approach I have outlined to issues about reliability is able to resolve some of the traditional difficulties that have been directed against reliabilist approaches to knowledge.[3] It is quite evident, however, that considerations of probabilities of success with respect to various kinds of judgments, while they may serve as some kind of standard against which our policies of expansion by incorporating information from others may be measured, rarely figure explicitly in our deliberations. Instead, we rely on rules of thumb, markers of others as having particular forms of expertise, that have been historically variable

[3] In particular, I think it can answer the very important challenge leveled by Robert Brandom in his *Articulating Reasons* (2000).

and that can be checked piecemeal for their ability to accord with the assessments that would result from a fully explicit canvassing of reliability. As we look into the past, we can recognize that some of the social characteristics used by our predecessors in guiding their strategies of incorporation were ill-adapted to the goal (theirs as well as ours) of acquiring error-free information: The seventeenth-century tendency to make social class an indicator of reliability strikes us as misplaced.[4] In accordance with the pragmatist themes presented in the first section, we can conceive our own epistemic task in terms of the improvement of the social markers and the social institutions that guide contemporary strategies of incorporation, a project that, because of the pervasive character of those institutions in all our inquiries, can be undertaken only piecemeal. The remainder of this chapter takes some first steps toward clarifying and carrying out that task.

III

The simple decision problem of the last section made it apparent that the value of the information potentially received makes a difference to the merits of incorporation – and this, of course, is a point from classical pragmatism that Levi has insisted on throughout his career. In the situation of face-to-face encounter – the simple mode of transmission that provides the basis for Craig's identification of the function of knowledge – there is little difficulty in locating the sort of value that is of concern. We focus on the value of the information for the potential recipient (for *X* in my schematic account). How the source evaluates that information is not relevant. It is simply a part of the background that others may be expected to issue a verdict with respect to a particular issue (recall that I made the simplifying assumption that posing the question would elicit some answer). Yet this inspires an obvious question: Why are there potential informants for the topic that concerns the putative beneficiary (*X*)? A plausible answer, given that people are finite cognitive agents and do not take on board items of information that they perceive as trivially worthless, is that these informants either assign some value to the information or else function as cognitive altruists who acquire the information because they recognize it as valuable to others. If this is correct, then we should think about the ways in which *differences* in assignments

[4] For this example (and other very interesting cases), see Steven Shapin, *A Social History of Truth* (1994). While I disagree with Shapin's attempt to couch his studies as contributions to the "social history of truth," that does not detract from the value of his work in elaborating the kinds of criteria used to identify those from whom our predecessors thought they could learn.

of value might undermine the possibilities of knowledge transmission, how greater harmony might be achieved or greater cognitive altruism fostered. These matters arise with far greater force as we shift our focus from the Craigian exchanges between individuals to the impersonal systems of public knowledge.[5]

Like Peirce and Dewey, Levi embeds his knowing subjects in communities of inquirers, but there is no fully explicit pragmatist treatment of the ways in which individuals interact with one another through the construction of public sources from which all can draw. The functionalist account of the last section is similarly one-sided, taking it as given that some are able to transmit what others need. Effectively, the discussion has begun the exploration of wise epistemic consumption without giving thought to the forms of epistemic production, an omission that is readily understandable when one conceives of the single individual building up a corpus of belief but becomes much more problematic once the individual is viewed in relation to others.

Indeed, the functionalist account so far developed is compatible with a laissez-faire picture of the economy of knowledge. There are all of us, individual inquirers, each assigning values to various kinds of information. We go about, acquiring information as we can, sometimes finding that we can save time and trouble because others have looked for the same kind of information, and when that happens, we use the considerations of reliability (or our proxies for them) to incorporate what they tell us or to withhold our assent. But the laissez-faire picture is surely inadequate both as a presentation of how knowledge-seeking actually works and as a view of how it ought to work to fulfill the epistemic aims of individual inquirers. Most of the information obtained from others is gained not by posing questions to individuals, but by consulting the sources that people collectively have constructed. From ancient times to the present, human beings have elaborated public systems of knowledge, with an eye to providing information that others might want. Further, we might think, that is a good thing, because this more cooperative project can be expected to benefit each of us.

Yet we cannot think of the public systems of knowledge by taking them to be the huge corpus of some massive individual and simply applying the picture Levi provides for individual belief change. No single function measures the

[5] In effect, people rarely face the simple decision problem of section II. Instead, they acquire from their social milieu rough-and-ready ideas about the relative reliability of various potential sources. If they are lucky, there will be sources for the information that concerns them, and they must then use their assessments of the value of information, the socially transmitted views about relative reliability, and their estimates of the cost of search to decide whether they should seek out sources and, if so, which ones.

way in which value is attributed to items of potential information; there is no uniform way of trading value of information against risk of error. Rather, the public system is built from the efforts of distinct individuals who value different kinds of information and who variously weigh the cost of error. The main features of its production should be made explicit.

Think of the public system as a depositary. Items of information can be placed in the depositary – inscribed on the books, as it were – provided that they pass the test of admission. The pertinent test may be applied to the items directly (focusing on the accompaniments with which they come), or to the depositors, or may involve some mixture; communities of inquirers may amass the deliverances of particular kinds of people (seers, gentlemen, people with doctoral degrees), or they may demand expressions of evidence (photographs, proofs), or they may use a criterion that mixes both factors (the seriousness with which a proposed item is evaluated may depend on the status of the person who transmits it). Those items that pass the approved test may be available for other members of the society, possibly for all of them, although the possibilities of access can easily be restricted either by design or simply because the information is formulated in ways that require conceptual sophistication.

I want to raise three general problems that attend the production of systems of this kind. First there is the trade-off made between the value of information and the disutility of error. In effect, there will be a community-wide commitment to treating some individuals as knowers, even though there is no single person who represents the community and performs the kind of assessment envisaged in the functional account of the last section. That commitment is effectively presupposed by those who draw from the public system, so that whatever propensity they may have to trade information against error-avoidance, some kinds of decision have already been made for them, affecting the ways in which they frame their own inquiries. Second, the system embodies decisions about the kinds of information worth acquiring, decisions that may variously represent the interests of those who deposited the information and that may be quite differently related to the needs and wants of those who later come to use it. Third, in organizing the information and in setting conditions for access to it, the community makes further decisions that can affect the ability of people to retrieve items that are relevant to their concerns.

Each of these sets of decisions involves balancing conflicting desiderata. To place the bar for admission to the stock of public knowledge very high would diminish the chance that what is inscribed on the books contains error, but, because less is inscribed, the project of building on the effort of past investigators may go more slowly. To think broadly and inclusively about accommodating the diverse interests of different individuals (or groups) is likely to slow the rate at which any particular mode of inquiry can be pursued and also, perhaps, to complicate issues of disseminating the information acquired. To present information as accurately and precisely as possible may well require modes of formulation that erect barriers against ready access.

There is a fourth source of conflict that may easily go unnoticed. The system of public knowledge is envisaged as a continuing project that, at any given stage, offers points of promising further development as well as opportunities for application to human needs. Often, these different kinds of ventures will be in competition, and when that occurs, the strategies for further development will incorporate a value judgment about how the balance is to be struck.

Indeed, judgments of value, not fully articulated by any individual, effectively underlie the public knowledge we have, written into the way in which those who produce knowledge are identified, into the types of information that are sought, into the responsiveness to potential users of the knowledge, and in the extent to which the continued development of public knowledge is given priority over – or sacrificed to – the delivery of solutions to immediately urgent problems. Levi's explorations in epistemology and the philosophy of science are noteworthy for their sensitivity to the intertwining of judgments of fact and of value. I am suggesting that a more explicit consideration of the system of public knowledge, a system that is enormously pervasive in the beliefs of all of us, would expose an even richer array of value judgments and of complex questions about how to improve the social arrangements for inquiry.

IV

I want to close by offering a general account of the social epistemological problem complex and integrating it within the functionalist approach to knowledge articulated in section II. I start with a classification of public resources for deliberating agents.

The first division is between those parts of the system that purport to deliver truth and those that aim at something else (the presentation of possibilities, for example). Within those that aim at truth, there are differences in the levels of exactness claimed. The community effectively makes an assignment of topics among these categories (possibly allowing that a particular question be treated at several levels of precision). It must also embody a decision about the conditions that contributions should meet, the organization of contributions, and the modes of access. For it is evident that, in a world brimming over with all sorts of potential information, it is impossible for people to acquire all the skills that would give them access to the most precise and accurate sources of information on all topics – hence the development of information providers at different levels of technicality (think of the spread from the specialist scientific journal through publications such as *Science* and *Nature* to *Scientific American*, popular science periodicals, and the science reporting of newspapers). All this is to serve a function (shortly to be characterized more carefully) of assisting individuals in their inquiries and deliberations.

Recognizing some parts of our public resources as not in the business of generating truth is intended to acknowledge the historical variability of the types of knowledge central to community inquiry. Plainly, there have been important differences among communities concerning the status of advice about how to live. Within many religious traditions, that sort of advice is perceived as making a claim to truth, and it is central to the conception of public knowledge. From the perspective of secular liberal democracy, however, this kind of advice must be reassigned: After Mill, we think of the descriptions of "experiments in living" as important resources for individuals to contemplate, but not as having the authority to prescribe how one should live.

It is valuable to have those descriptions because they can serve deliberators in planning their lives and because they may frame the ways in which they conduct their inquiries. Should some of them be given a special kind of authority, being seen as such vital and important possibilities that they are especially worthy of consideration? The question leads quickly to issues about the desirability of canonical texts and (assuming that there are to be some) to the scope and contents of the canon.

My principal interest here, however, lies with the parts of the public system that lays claim to truth: the knowledge business. Its function is to provide a source of true claims on which people, with their diverse interests and projects, can draw. But which people are to count? And how are we to understand their interests? Both questions lead into deep and difficult problems.

For a simple individualist pragmatism, the values assigned to particular types of information are simply given: These are the sorts of things the subject wants to know, and we are to think about routines for acquiring the pertinent answers. One might adopt the same perspective for public knowledge, seeing it as answerable to the public's actual desires for relief from ignorance. Yet, in many cases, the desire itself may be premised on ignorance – people want to find out about this and have no concern with discovering that precisely because they don't know things that others do. Part of the function of the public system of knowledge, one might suggest, is to help guide people's wishes and values (that, after all, is what makes plausible the inclusion of resources not aimed at truth, descriptions of "experiments in living"). A public system of knowledge that fails to identify the exact place in modern Turkey on which Noah's ark came to rest is not at fault for failing to deliver what a hefty percentage of Americans are burning to know, for it contains other items of information that, if fully absorbed, would tell decisively against the presuppositions of the question. The epistemic welfare of individuals is to be assessed not in terms of answering questions to which they give priority but in relation to questions that they would see as important if they absorbed parts of the available public knowledge of which they are currently ignorant. (Working out the details of an account of epistemic welfare is obviously extremely hard, and I intend my remarks here only as motivation for undertaking the project.)

Even if we were to have a fully worked-out conception of individual epistemic welfare, there would be serious questions about how to characterize the function of the public system, for we would have to decide how to represent the *collective* epistemic welfare. Who is to count? And with respect to those who count, can the collective epistemic welfare simply be viewed as the sum of the individual contributions? With respect to the first question I am inclined to defend an inclusive answer, not simply thinking of those who stand in some special relation to the principal contributors to the public knowledge system: It would be wrong, I think, to suggest that co-nationals of those who advance our knowledge are especially privileged to have their interests represented in the problems addressed. This judgment is consequential, for it entails that the public system of knowledge we now have is wrongly biased toward the needs of affluent people.[6] That conclusion will be underlined if one answers the second question by suggesting that collective epistemic welfare is increased more by those endeavors that advance the epistemic welfare of those who are currently worst off.

Some of the difficulties that arise in articulating and defending answers to these questions may be avoidable if we conceive the project of social epistemology in accordance with the pragmatist themes with which I began. Suppose that we have some rough understanding of individual epistemic welfare and focus on those changes that would be Pareto improvements, making some better off while leaving nobody worse off. Even without a precise account of individual epistemic welfare, we may be able to recognize that some modifications of our current practice would be superior (or inferior) according to this criterion.

With these considerations as background, I want to close by contrasting the epistemic picture I have been sketching with that presented by Isaac Levi.

[6] See James Flory and Philip Kitcher, "Global Health and the Scientific Research Agenda," *Philosophy and Public Affairs* 32, no. 1 (2004): 36–65.

Let me start with a minimal characterization of a public system of knowledge: A system of this kind exists within a community *C* when there is a practice of identifying some propositions as items of knowledge or as identifying certain individuals as being knowledgeable with respect to particular kinds of propositions, a practice that is shared among members of *C* and typically used in training the children born to members of *C* (as well as being available to those children when grown up, as the source of further information). What advantage would members of the community have with respect to cognitive agents who operated completely independently or who depended on one another in the manner envisaged in section II?

I have already noted the obvious cost of proceeding entirely independently: You just don't acquire very much information. But reflection on the decision problem of section II should make it apparent that simple face-to-face borrowing is unlikely to do much better. To try to assess the reliability of others without making use of social markers would require a lengthy process of monitoring performance under conditions where one was also in an independent position to evaluate the information supplied. In effect, duplication of effort serves as a preliminary calibration so that other people can be viewed as epistemic extensions of oneself. Provided that the social markers correlate fairly well with the reliability of the agents they pick out, use of them will greatly enlarge the range of individuals from whom any of us can borrow. It is plausible to hold that, under a system of public knowledge, each of us can expect to be at least as well off epistemically as we would otherwise have been.

But there's an obvious worry. Even if the practice of the community works well on the whole, its identification of some doctrines or people as authoritative may enshrine some pervasive errors, which serve as sources of infection in the community's epistemic life (and beyond). Hence arises the obvious thought that the traditional wisdom must be subject to criticism. Encouraging criticism of everything at once would be to return to the presocial state. Moreover, as Levi has persuasively argued, one needs to have reasons to open one's mind: Doubt should be genuine (as Peirce so eloquently insisted). From the community perspective, then, it is good that there should be people who occasionally play the apparently thankless role of exploring whether there are grounds for questioning what has long been settled. As I have suggested elsewhere, the existence of mavericks is encouraged by the reward system that governs knowledge production; to follow the existing rules can assure you a satisfying and comfortable career, but to win the largest prizes, to become recognized in the history of human knowledge as a great figure, requires successfully overturning what has previously seemed secure.[7]

Effectively, then, the community serves as a bureau of standards in matters of knowledge, as it does in other areas, and the existence of that bureau looks like a good thing. The extent to which propositions, individuals, methods, or systems of training are marked as authoritative is, however, open to question. Some may hold that individual choice with respect to the modification of belief should be as wide as possible. Once individuals have been given the basic equipment with which to appraise others, they can be left to make their own decisions about which sources are worthy of their trust. The epistemic analogues of market forces will yield a satisfactory outcome.

But there are serious reasons for doubt. Economists have recognized that asymmetries in information subvert the operation of markets, so there are general grounds for wondering whether potential consumers of information will be well served by a system in which those from whom they might draw have an interest in appealing to large numbers.[8] When the dissemination of information becomes a business, as it plainly has, the absence or irrelevance of any scheme for accrediting sources will give rise to a "free market" in which there is no reason to expect a particularly good outcome. The result may be a situation in which millions of people are catastrophically misinformed about issues that matter to them – believing, for example, that a terrorist attack involved citizens of a nation with which a dishonest government wants to wage war.

The main thrust of my argument has been that Isaac Levi's approach to the growth of knowledge needs to be extended to recognize the intricate ways in which social factors affect the modification of our beliefs. In the last paragraphs, I have been guarding against the possible objection that one might easily return to the preferred Levite state, either by repudiating public systems of knowledge or by limiting their extent. (I do not mean to suggest that Levi himself would advocate either of these suggestions; his writings make no commitment on the topic, but the suggestions would defuse my charge that his epistemology is incomplete.) If we take the deepest theme in pragmatism seriously, supposing that philosophical problems arise out of human needs that evolve with our societies, then, I maintain, we should neither remain obsessed with the legacy of problems we have inherited from Descartes nor

[7] See *The Advancement of Science* (New York: Oxford University Press, 1993), chap. 8.
[8] For incisive discussions of market behavior under asymmetries in information, see G. A. Akerlof, *An Economist's Book of Tales* (1984).

rest content with the valuable transformation of those issues into questions of methodology, begun by Peirce and continued by Levi.

We need to confront the full messy reality of the ways in which knowledge today is socially produced and socially distributed, not with the debunking attitude sometimes found among historians and sociologists of science, but with the constructive search for improvements that marks the pragmatist tradition. The abstract descriptions I have used above, in my references to items deposited for potential consumers who may have access to them, need to be given concrete and detailed form, so that we can begin to understand what we are doing and how we might do it better. We need to understand the requirements that govern the formation and maintenance of databases, typically put together in historical reality by agents with different purposes and with different propensities to trade informational value against the risk of error. Unlike individual agents, the foci of Levi's explorations, these databases cannot be held to simple standards of consistency. Further, as may already have become apparent, epistemology so broadened becomes connected to a range of questions that occupied Dewey, questions about the intertwining of values with claims to knowledge, about the role of education, and about the role of information in democracy. Epistemology ought to be responsive to the problems about knowledge that arise in our time, in evaluating the social institutions that serve as resources for democratic societies, and not simply confine itself to the issues that have arisen from the historical predicaments of the past. Peirce's reaction to Cartesianism sounded this theme a century ago, and I believe we need to undertake a similar shift in the philosophical agenda today.

Dewey's own attempts to grapple with knowledge broadly construed are often imprecise and so, I must confess, are mine. The challenge is to bring to this wider perspective the kinds of precision that Peirce began and Isaac Levi has continued in the case of individualistic epistemology. But that would be a task for a *vastly* longer inquiry.

4

Infallibility and Incorrigibility

Bengt Hansson

In his book *The Enterprise of Knowledge*, Isaac Levi makes a distinction between infallibility and incorrigibility and remarks that many authors, among them Peirce, do not seem to be aware of the difference. Levi's distinction is made in the context of his discussion of fully justified belief, or "knowledge as a standard for serious possibility."

To be (epistemologically) infallible about a piece of knowledge is to ascribe maximum certainty to it, in the sense that anything inconsistent with it will not be regarded as a serious possibility. According to Levi, we are entitled to claim infallibilism for a certain corpus of knowledge that extends beyond logical, analytical, mathematical, and other a priori truths.

To claim incorrigibility for a piece of knowledge is to claim that we will never change our mind about it, that it will still belong to our corpus of knowledge no matter what new evidence will come across. It is permanent, rather than maximum certainty.

Levi's standpoint is that although we are entitled to disregard certain possibilities as not being serious (and hence to claim epistemological infallibility about their negations), we must still allow the possibility that we will change our mind in the future, that is, we must not claim incorrigibility.[1]

In the latter respect, he agrees with common morality, which, at least in scientific contexts, is to be prudent about one's conclusions and to admit the possibility of being mistaken and therefore to admit corrigibility about anything that is not a priori. What confuses things is that the term used for this attitude is sometimes "fallibility."

For Levi, to claim incorrigibility is stronger than to claim infallibility (a view that I challenge below), and he is therefore consistent in claiming the latter without the former. But in claiming the latter, he finds himself at odds

[1] Like Frege, Levi keeps logical and psychological concepts strictly apart. His objections against incorrigibility are purely logical. He has, to the best of my knowledge, never argued against psychological incorrigibility (i.e., what laymen call stubbornness).

with Peirce's pronounced fallibilism, which, besides the major fault of being at odds with Levi's own view, also has the minor fault of being apparently paradoxical in itself.

For it seems that Peirce's fallibilism entails incorrigibility, for roughly the following reasons: Fallibility means that one should admit the possibility of mistakes and therefore claim full knowledge only in exceptional cases. In the extreme, the only exceptional cases are logical truths. But these will not change, come what may. So if we are *very* reluctant to claim full knowledge, then the things we *do* know will become incorrigible![2]

In this chapter, I wish to defend two general theses that are relevant to these questions. First, that fallibility and corrigibility are independent notions, and second, that infallible pieces of knowledge should not be identified with those that are maximally certain.

My position has been formed during my work on a larger project about the nature and representation of knowledge. In that work I have become convinced that one cannot study full-fledged knowledge without also discussing partial knowledge or belief. I regard full knowledge as a hyperplane in the space of epistemic states, or, metaphorically, as the upper surface of the deep sea of more or less justified opinions. As in the real sea, most phenomena observed on the surface are best explained by processes operating beneath it.

In this space of epistemic states, I distinguish between two "dimensions": certainty and robustness. To attain maximum certainty is to reach the surface of the epistemic sea, that is, to have full knowledge. Robustness is inertia to revision, or insensitivity to new information. Maximum robustness is incorrigibility.

My strategy is to argue, first, that certainty and robustness are distinct and independent concepts for epistemic states in general, and, second, that they remain distinct also in the subspaces where they attain their maximum values.

I make no attempt at a formal definition of robustness but rather explain it by way of examples and by contrasting it with probability. We may say that the task of probability is to reflect one aspect of the incompleteness of knowledge, the certainty aspect. But the probability value in no way exhausts the dimensions of incompleteness of knowledge. Items with the same probability may, for example, react quite differently to new information.

Consider an interviewer who is polling the general opinion about a certain issue. He may make an educated guess already after, say, thirty interviews, but

[2] The correct solution to Peirce's paradox is, of course, that this particular kind of incorrigibility is harmless, which is not to say that its presupposition, extreme fallibilism, is commendable.

his estimate will be much better underpinned when he has made all his 1,500 planned interviews. The latter estimate is much more robust than the first guess – any kind of additional information, for example, five more interviews with deviating answers, would affect the guess much more than the final estimate. This is true even if the final estimate should be numerically identical to the first guess.

It is enough for my purposes to establish robustness at the conceptual level as a sort of inertia to revision and distinct from probability. Perhaps it is not numerically measurable, perhaps it is not even a complete ordering, but it can be defined very strictly, as the above example shows, as a partial ordering, and it is reasonable to assume that this partial ordering can be extended to a useful concept, which may be, however, more vague.

Certainty and robustness thus span independent dimensions in the space of epistemic states, and we proceed to look at the maximum values of each. Maximum certainty, which must be more than unit probability, is (at least roughly) infallibility. Maximum robustness is when further inquiry is pointless, which is to say that one's epistemic position is incorrigible.

The question of whether these maximum values can really be attained is an important one that requires separate treatment. I touch slightly on it below, but a thorough discussion would be too extensive a digression in the present context. Let it suffice to note that Levi on some occasions uses the phrase "maximum certainty" for such knowledge as is standard for serious possibility. There is no need to treat robustness otherwise; maximum robustness means that no further inquiry is meaningful, and we may think about it as the "standard for serious inquiry."

These two maximum concepts are independent in the space of epistemic states in general. That maximum certainty (infallibility) does not entail maximum robustness (incorrigibility) is Levi's thesis, and I agree, so I say no more about that.

Nor does maximum robustness (incorrigibility) entail maximum certainty (infallibility). My epistemic status concerning the possible disintegration of this U-238 atom during the next hour will not improve once I have ascertained that it is really a U-238 atom. And the chance that a red ball will be drawn from this urn in the next fair drawing cannot be better known than it is when I have inspected all the balls and noted their color and that they are equal in all aspects but color.

Given that we deal with epistemic states in general, the two notions are easily kept apart, logically as well as psychologically, but the question of the fine structure within the hyperplane of maximum certainty still remains to be discussed. When Levi speaks about incorrigibility, he sometimes uses the phrase "permanent certainty," as in the following quotation about the interpretation of Peirce's fallibilism: "Peirce, however, seems to have intended to claim more than that we cannot attain permanent certainty concerning any matter of fact. He meant to deny that we can attain *maximum* certainty."[3]

This can be interpreted in two ways. Either "permanent certainty" means incorrigibility. Then Levi in effect says that incorrigibility is stronger than infallibility (maximum certainty), thus denying their independence. Or the word "permanent" signifies by itself incorrigibility, and "permanent certainty" refers to the more restricted notion of incorrigibility within the realm of the (maximally) certain. Then it is analytical that permanent certainty is at least as strong as infallibility.

Obviously, in view of my previous argument, I reject the first interpretation. But the second interpretation is somewhat awkward to combine with the view that we are entitled to infallibility but not to incorrigibility. To see why, let us first assume that the corpus of infallible truths consists of those truths that are maximally certain in a strict sense. That means that the relation "more certain than" is coarse-grained enough not to make any further distinctions between the various elements of this corpus. But if the same elements are ordered by the relation "more robust than," there must be no maximum elements, or these would incorrigible. So, this relation must be fine-grained and on an infinite set. There is no logical conflict involved in the two relations failing to be parallel, but it is hard to see what reasonable evidence there could exist for this being in fact the case.

It seems more promising to assume that the talk of "maximum certainty" is a figure of speech that should not be taken literally but interpreted to mean "certain enough to exclude incompatible options from being serious possibilities." This view gains support from Levi's own examples of things that are not serious possibilities. He mentions a coin flying out toward Alpha Centauri when tossed. The negation of this therefore belongs to the person's corpus of full knowledge and is infallible in Levi's sense. But the coin flying out toward Alpha Centauri with a speed approaching that of light is an even less serious possibility (because it is more specific), and its negation is therefore even more certain. It is thus possible to make finer distinctions between degrees of certainty within a corpus than its status of infallibility requires.[4]

But the same trick could with equal right be applied to robustness. Instead of maximum robustness, we may speak about being "robust enough to make

[3] *The Enterprise of Knowledge* (Cambridge, Mass.: MIT Press, 1980), p. 14.

[4] It is not necessary that these degrees be measurable by real numbers. It is not even necessary that they form a complete ordering.

it pointless to pursue inquiry any further" without denying that more fine-grained distinctions can be made in principle. Then the claim that even infallible knowledge is corrigible becomes the claim that nothing sufficiently certain is also sufficiently robust (disregarding logical and analytical truths). This is a priori implausible, given our previous examples. And a positive counterexample would be my knowledge that this Pb-206 atom will not disintegrate spontaneously. It is both sufficiently infallible and sufficiently incorrigible.

So I find considerable problems with upholding the view that infallibility and incorrigibility should be treated differently. The most plausible approach seems to be to admit that very fine-grained distinctions can be made with regard to both certainty and robustness and that we may select various cut-off points for various purposes or, like Peirce, shun cut-offs altogether for theoretical purposes. This is not to yield to general vagueness or indeterminateness, because consistency conditions between various choices may apply and be very strict.

I have not addressed this issue because I consider it to be such an important problem in itself, but because it is an example of a widespread and problematic tendency, namely, the neglect to observe the dynamic aspects of knowledge. For I wish to emphasize not only that certainty and robustness are independent notions, but also that certainty is essentially static, or descriptive of a given epistemic state, while robustness is dynamic. Unlike certainty, it belongs to a family of concepts that are central to the theory of how epistemic states react to various types of input, the theory directly concerned with the pursuit of inquiry.

5

Why Inconsistency Is Not Hell

Making Room for Inconsistency in Science

Otávio Bueno

1. INTRODUCTION

On most accounts of belief change, inconsistent belief systems are an "epistemic hell" to be avoided at all costs (see, e.g., Gärdenfors 1988, p. 51). From a *normative* point of view, we can perhaps understand why this is the case. The underlying logic of most theories of belief change is classical, and classical logic is explosive, that is, everything follows from a contradiction. And a belief system from which everything follows should definitely be avoided. It is certainly of not much use if one wants to determine what one should believe and what one should do.

In a number of works, Issac Levi challenged this way of approaching the issue. On his view, there are contexts in which inconsistent belief systems are bound to happen. This is the case, for example, of observations. According to Levi, in some contexts it is legitimate to add a doxastic proposition to a belief system with which it is inconsistent: "Making observations and coming to fully believe propositions incompatible with one's initial convictions is a case in point" (Levi 1991, p. 68). The idea is that we may inadvertently tumble into inconsistency as the result of "deploying a reliable program for routine expansion" (ibid., p. 110), that is, as the result of adding a new belief to our belief system. In other words, *descriptively* at least, inconsistent belief systems simply happen, and this fact needs to be accommodated.

The trouble, however, is that despite acknowledging that expansion into inconsistency may sometimes be legitimate, Levi immediately adds that "it is always urgent to contract from an inconsistent state of full belief. The

My thanks go to Mark Colyvan for extremely helpful comments on an earlier version of this work.

contraction will remove either **A**, ∼**A**, or both" (ibid., p. 68). This urgency clearly derives from the trivialization (in classical logic) of the belief system resulting from its inconsistency. And for this very reason, Levi insists that an inconsistent belief system is unacceptable, since it "fails as a standard for serious possibility for the purpose of subsequent inquiry and for practical deliberation" (ibid., pp. 76–7; see also Levi 1996).

Levi has certainly gone a long way toward devising a broader, more encompassing approach to belief change with regard to inconsistency. And he can only be commended for that.[1] In this chapter, I argue that we *can*, and *should*, go further. It is not only that *descriptively* we need to make room for inconsistent belief systems – they are indeed a striking fact of our epistemic life. But also *normatively* to maximize informativeness or, at least, to minimize information loss, we often need to entertain and pursue inconsistent belief systems. To pursue inconsistent systems is a useful device for a number of reasons: (1) This is often the only way to explore inconsistent information without *arbitrarily rejecting precious data*. (2) Pursuing inconsistent systems is sometimes the only way to *obtain new information* (particularly information that conflicts with deeply entrenched theories). As a result, (3) pursuing inconsistent belief systems allows us to make better informed decisions regarding which bits of information to accept or reject in the end.

After motivating and illustrating each of these three benefits of inconsistent belief systems, with examples from the foundations of mathematics and the empirical sciences, I argue that Levi's own approach to belief change could become stronger by making room for inconsistency, without giving up the crucial features of his pragmatism. The crucial move is to change the underlying logic to a paraconsistent one, which, in consistent contexts, yields exactly the same results as the classical approach (see, e.g., da Costa and Bueno 1998). Once there is room for inconsistency, there is also room for informativeness in inconsistent belief systems. As a result, the emerging account accommodates both descriptively and normatively the

[1] See, in particular, Levi's recent response to Olsson's searching critique of his account of contraction from inconsistent belief states (Levi 2003 and Olsson 2003). If inconsistency is epistemic hell, Olsson asks how could it ever be rational to enter such a state, and how could one regain consistency? Levi resists the critique by noting that one could devise a routine "prior to inadvertent expansion into inconsistency when the deliberating agent embraces a consistent point of view" (Levi 2003, p. 141). This strategy, however, doesn't succeed in making room for inconsistency, given that it ultimately shifts the burden to the reliability of the presumed consistent belief systems prior to the expansion into inconsistency. An alternative approach is in order.

significant role that inconsistency plays in inquiry, without yielding an epistemic hell.[2]

2. BENEFITS AND COSTS OF INCONSISTENT BELIEF SYSTEMS

There are many reasons why we should take seriously inconsistent belief systems – and many *benefits* emerge from doing that, which I examine below. But there are also important reasons why we need to be *extremely careful* when dealing with such systems. So, before examining the benefits of inconsistent systems, it's important to be clear about the alleged difficulties – and hence the *costs* – posed by inconsistency, and determine whether, and to what extent, they are reasonable.

2.1. Costs of Inconsistent Belief Systems

Three major costs are typically associated with inconsistency:

(i) *Triviality*. As we saw above, a common charge against inconsistent systems is that they are trivial (given classical logic); that is, everything follows from them. So, inconsistent systems (whose underlying logic is classical) are useless as epistemic guides and as the basis for practical deliberation (particularly, given the goal of maximizing true beliefs and minimizing false ones). This is, of course, a major complaint, and ultimately it's what has given inconsistency such a bad name.

(ii) *Unreliability*. A further reason why inconsistent systems are not considered epistemically significant emerges from the fact that the world itself is not taken to be "inconsistent" – in the sense that there cannot be true inconsistent descriptions of the world. Thus, if such descriptions cannot be true – or, at least, not *completely* true – there are no grounds to think that inconsistent theories are even reliable. After all, if *no* inconsistent theory of the world *can be true*, no inconsistent theory *is* true, even in the long run.

Furthermore, it's also unclear how we could approach true, consistent theories via inconsistent ones. After all, if inconsistent theories are trivial (in classical logic), there's no guidance as to which parts of an inconsistent theory should be preserved and developed further, and which should be rejected. Nothing in such a theory would decide that. Note that the

[2] A note about the terminology: In this chapter, I use "inconsistency" and "contradiction" more or less interchangeably. Moreover, in the discussion of belief systems and belief change below, I also move more or less freely from talk of beliefs to talk of propositions that express such beliefs. Nothing of substance for the present discussion hangs on this.

situation here is substantially different from the familiar cases raised by Duhem and Quine. Their point is that logic cannot decide which part of an empirically false theory should be rejected and which part should be further developed. But in the case of a *consistent* theory, one can in principle still draw nontrivial consequences from the "package" of the theory under test plus auxiliary assumptions, additional theories regarding the measuring devices, and so on. By exploring additional consequences of the overall "package," one can, at least in principle, try to determine what's the best way to proceed. However, the case of an inconsistent theory (whose underlying logic is classical) is importantly different. Given the triviality of the theory (in classical logic), there's no way of exploring further consequences from the theory in any informative way: *Everything* follows from the theory. Under these circumstances, inconsistent theories are not of much use.

Moreover, if we take the aim of inquiry to be the production of *true, maximally consistent* belief systems (Levi 1991), then *inconsistent* belief systems are not an obvious starting point. In turn, if we start from *maximally consistent* belief systems, we may end up developing a *true* maximally consistent system in the long run. Given their lack of reliability, inconsistent systems are simply not the way to go.

(iii) *Lack of informativeness*. The charge of lack of informativeness of inconsistent belief systems is connected with the triviality complaint. If everything follows from an inconsistent system (in classical logic), such a system would be completely uninformative. We would be unable to use such a system to distinguish between bits of information that we have reason to believe – given the available evidence – and those that we haven't. After all, in classical logic, an inconsistent system would seem to give us equal reason to believe in everything – definitely not a welcoming result!

But do we have good reasons to accept these complaints? As will become clear, I don't think so. In fact, as long as we adopt the right framework to conceptualize and accommodate inconsistency, the alleged *costs* of inconsistent belief systems just discussed turn out to be significant *benefits*.

2.2. Benefits of Inconsistent Belief Systems

If we take seriously the examination of inconsistent belief systems, interesting benefits emerge. But note that to take such systems seriously, it's not required to take them as *true*. There are many possible attitudes one can have toward a belief system besides truth. A system can be accepted as *empirically adequate* only, as *quasi-true*, or as *warrantedly assertible*. We can simply *pursue the* *system* for the purposes of epistemic inquiry, or we can explore some of its consequences while *refusing* at least temporarily *to entertain* others, or we can take the system as just a *tool* for predictions about only a given range of phenomena. The point here is that one need not be a realist about inconsistent belief systems to take them seriously. In fact, the view I favor is thoroughly antirealist, and according to it, inconsistent theories need not be (and typically aren't) true; such systems are, at best, *quasi-true* – that is, roughly speaking, true given the partial information about a given domain and possible ways of extending that partial information (see section 3 below for details).[3]

(i) *Nontriviality*. The charge of triviality for inconsistent belief systems emerges only if the underlying logic of the system is classical. After all, classical logic is "explosive," in that every sentence in the language can be derived from a contradiction (i.e., a sentence of the form '$A \wedge \neg A$'). Although several logics are explosive in this sense (such as classical and intuitionistic logics), *not all* of them are. As is well known, paraconsistent logics are *not* explosive, and so, in these logics, contradictions do not entail everything (see, e.g., da Costa 1974; Priest 1987; Priest, Routley, and Norman 1989; and Carnielli and Marcos 2002). As a result, as long as the underlying logic is paraconsistent, the triviality charge does not arise.

(ii) *Reliability*. With an underlying paraconsistent logic, it's not difficult to respond to the lack of reliability worry. First, the reason why the world is not taken to be "inconsistent" – in the sense that no inconsistent description of the world is taken to be true – is that classical logic simply rules out that possibility. After all, in classical logic, contradictions are necessarily false. But in some interpretations of paraconsistent logic, such as Priest's dialetheism (see Priest 1987), some contradictions are true. This is the case, for example, of the liar sentence ('This sentence is not true') that for familiar reasons comes out both true and false. This opens the possibility that, at least in principle, there might be true contradictions about the world. In which case, inconsistent theories need not be unreliable.

Even though I don't think we should consider an inconsistent belief system to be true, this is *not* because the information provided by such a system is necessarily false. The existence of paraconsistent approaches establishes how

[3] The notion of quasi-truth was formulated and applied to a variety of cases by Newton da Costa and Steven French in several works. They defend a *realist* interpretation of that notion, but as they note, an antirealist reading is also possible (for details, see da Costa and French 2003 and the references quoted there; see also section 3, below).

inconsistent theories *could be* true, and this is done independent of Priest's more radical interpretation that insists on the *existence* of *true* contradictions. Rather, as part of an overall empiricist approach to inquiry, if scientific theories need not be true to be good, in particular, *inconsistent* scientific theories need not be true either. Nothing in scientific practice *requires* a commitment to the truth of the theories in question.[4] And hence, nothing in such a practice demands a commitment to the *truth* of *inconsistent* theories.

Instead of simply outright banning inconsistent theories, we could use them to generate better, consistent successors, by exploring, in a paraconsistent setting, the consequences that inconsistent theories have. Only after such an exploration can we identify the features of the theories in question that we may have good reason to preserve. In this sense, properly explored, inconsistent theories may be reliable or, at least, could be used to generate better theories in the long run.

Note that throughout this process, the aim of inquiry can still be truth (or perhaps quasi-truth). But there's no reason to assume that a maximally *consistent* belief system is a better guide to truth than a maximally *nontrivial* belief system (see, e.g., da Costa 1997). In fact, once the distinction between inconsistency and triviality is drawn in a paraconsistent setting, a maximally *nontrivial* system (even though inconsistent) may be a better guide for the generation of a *true* maximally consistent system than a maximally consistent system alone. After all, with more information to explore in an inconsistent setting, better theories could be formulated.

(iii) *Informativeness*. Once the triviality charge is accommodated, the complaint regarding informativeness can also be dealt with. Inconsistent belief systems need not be uninformative. After all, given that in a paraconsistent setting not every sentence follows from a contradiction, an inconsistent belief system won't provide us with (equal) reason to believe in everything. Some – but *not all* – sentences will be entailed by other sentences

[4] According to Jody Azzouni, the commitment to the truth of a scientific theory is indispensable, given that often we need to endorse a theory that, for a variety of reasons, cannot be explicitly stated (Azzouni 2004). Even if we grant him the point, it doesn't follow that the resulting notion of truth is substantive in any way (as Azzouni himself acknowledges). All we need in this case is a deflationary account of truth, which is essentially a logical device that allows us to assent to sentences in blind contexts (i.e., contexts in which the content of the relevant sentences has not been specified). But the commitments that a deflationary notion of truth brings (if any) need not be problematic at all (see Azzouni 2004).

in the system. But, to insist, there will always be sentences that are *not* so entailed. In this way, in a paraconsistent setting, the support given to the various beliefs of an inconsistent belief system is not uniform. There's genuine information to be explored, and just as with a consistent system, some sentences are better supported and more entrenched than others. Hence, in a paraconsistent setting, typically even inconsistent belief systems can be informative.[5]

There are three additional benefits related to the informativeness of inconsistent belief systems that are worth noting:

(a) In a classical setting (one in which the underlying logic is classical), faced with an inconsistency, one is committed to reject more or less a priori some data – perhaps even valuable data.[6] Exploring, in a nontrivial way, inconsistent belief systems provides an alternative way to *investigate inconsistent information* without arbitrarily rejecting available information. This provides a better mechanism to revise the original inconsistent system. After all, in a paraconsistent setting, it's possible to make inferences nontrivially in the presence of inconsistencies. As a result, we would be in a better position to devise improved, consistent formulations of the original inconsistent system, without losing significant information along the way.

(b) Pursuing inconsistent systems is sometimes the *only* way to obtain new information, particularly information that conflicts with deeply entrenched theories (see Lakatos 1978 and Feyerabend 1988). For example, when Bohr developed his atomic model, he articulated an inconsistent proposal, given the accepted theories at the time. On the conception of the atom he adopted, an atom was thought of as a minuscule planetary system, with electrons orbiting round a positive nucleus. But, according to electromagnetism, an atom of that sort would be radically unstable and would almost immediately collapse.

[5] Of course, this will depend on the paraconsistent logic and the contradiction in question. After all, if only one sentence is not entailed by the contradiction, the logic is paraconsistent, but, in such a case, the resulting theory is still uninformative. However, the paraconsistent logics I have in mind here, such as da Costa's *C*-logics (da Costa 1974), are more discriminative with regard to explosion than that.

[6] It might be argued that, in the presence of an inconsistency, the procedure of rejecting some information is perfectly justified. Given that, assuming classical logic, no inconsistent belief system can be true, at least some information in that system must be false, and hence part of the system must be rejected – or so the argument goes. The trouble, however, is to determine which bits of information to reject, in a way that both preserves as much information as possible and doesn't rule out bits of information arbitrarily. Much work in belief revision attempts to model this process. But being able to reason with inconsistent information provides an alternative pathway for that – including alternative strategies to obtain new, consistent versions of the original inconsistent system.

Bohr essentially ignored the inconsistency, introducing by fiat a postulate to the effect that

> [e]nergy radiation [within the atom] is not emitted (or absorbed) in the continuous way assumed in the ordinary electrodynamics, but only during the passing of the systems between different "stationary" states. (Bohr 1913, p. 874)

Despite the inconsistency, Bohr's model was *not* trivial. Bohr certainly *didn't* derive everything from his model, but managed to obtain the correct predictions – for example, regarding the wavelengths of hydrogen's line emission spectrum, among other items. Developing the model at the time was the *only* way to obtain such results, despite the inconsistency with accepted theories.[7]

A similar phenomenon is also found in nonempirical sciences, such as mathematics. For instance, as is well known, Frege developed a logicist reconstruction of arithmetic, but, inadvertently, in an inconsistent setting (Frege 1950, 1962). Significantly enough, though, Frege *didn't* derive everything from his inconsistent principles, but managed to show how arithmetic could be reconstructed from second-order logic and some definitions. Moreover, at the time, Frege's reconstruction was the *only* way of obtaining the logicist result. In fact, Russell's theory of types hadn't been developed yet, and even if it had been, it's unclear whether that theory meets the logicist requirements introduced by Frege.

Recent attempts to save Frege from contradiction – for example, by jettisoning the inconsistent Basic Law V and taking Hume's Principle as basic[8] – have the benefit of restoring the consistency of Frege's approach (see Wright 1983; Boolos 1998; and Hale and Wright 2001). But the resulting proposal has the cost of introducing a system that is not as obviously logicist as Frege's. After all, it is at least a *contentious* issue whether Hume's Principle is analytic, and so whether it could be legitimately considered an adequate basic principle for a logicist reconstruction of arithmetic.

[7] Lakatos has a fascinating, although somewhat idiosyncratic, discussion of this episode (see Lakatos 1978, pp. 55–68).

[8] Roughly speaking, Basic Law V states that every concept has an extension (of the objects that fall under that concept). The only essential use of Basic Law V made by Frege was to obtain Hume's Principle. According to that principle, two concepts are equinumerous if and only if there is a one-to-one correspondence between them. And using the latter principle, plus second-order logic and some definitions, Frege managed to provide a logicist reconstruction of arithmetic (Frege 1950, 1962). So, the idea was to reject the inconsistent Basic Law V, adopt Hume's Principle as the basic principle for the logicist reconstruction of arithmetic, and carry out Frege's approach from there (see Wright 1983; Boolos 1998; and Hale and Wright 2001).

(c) The considerations above indicate that, properly conceptualized, inconsistent belief systems, whether about empirical or nonempirical domains, can provide invaluable information. Even though an inconsistent system may only be a first, but important, step toward a consistent successor, the point still remains that such a system is significant. As we saw, in a paraconsistent setting, it's possible to take the inconsistent information at face value and explore its consequences, thus obtaining more information about the inconsistent domain (see da Costa and Bueno 2001). As a result, in the presence of an inconsistency, instead of simply ruling out some information blindly and a priori, we are in position to make better informed decisions regarding which bits of information to accept or reject in the end. The importance of information is, thus, maximized, while minimizing informational loss. And this is achieved *not* by simply accepting a contradiction and resisting any pressure to revise the system. Contradictions *are* problematic – some of them may trivialize a belief system even in the presence of a paraconsistent logic (see, e.g., da Costa 1974). Rather, in a paraconsistent setting, informational loss is minimized given that the consequences of an inconsistency are explored without ruling out a priori extant information. As a result, inconsistent information can be taken at face value, which allows us, in turn, to examine in creative ways the domain of the inconsistent but nontrivial.

It might be complained that this same result (namely, the exploration of consequences of inconsistent belief systems) can be obtained simply by using classical logic. After all, faced with a contradiction of the form $A \wedge \neg A$, we can always reject the contradiction, assume separately each side of the conjunction, and explore the resulting consequences of each assumption to find out additional information about the respective domains (one in which A is the case, the other in which its negation is the case). Then, by comparing the outcomes of each option, it's possible to decide which bits of information eventually to reject and which to accept.

This strategy has several problems. First, it eventually amounts to the rejection of some bits of information, without ever taking the conjunction that constitutes the contradiction in question at face value. Of course, the latter strategy can be implemented only in a paraconsistent setting. Second, if it's stated that there are good reasons to reject the contradiction (given that the latter is necessarily false), one is simply begging the question against the paraconsistent proposal. Finally, not all inconsistent belief systems have the simple presentation assumed in the strategy above. For example, in Frege's system, the contradiction is "buried," as it were, in a particular instance of Basic Law V.

Thus, as long as the appropriate logic is in place, there's no reason to think that inconsistent belief systems are trivial, unreliable, and uninformative. In the end, there are clear benefits associated with such systems.

3. MAKING ROOM FOR INCONSISTENCY: PARTIAL TRUTH AND PARACONSISTENCY

Having *motivated* some of the benefits of inconsistent belief systems, how can such benefits be *achieved*? What we need is a framework that (1) allows us to represent formally the idea that belief systems (in particular, scientific theories) need not be true to be good, and (2) accommodates inconsistent belief systems (in the sense that the latter are not trivialized in the presence of contradictions). To achieve (1), the notion of partial (or quasi-) truth is introduced. To achieve (2), an underlying paraconsistent logic is, of course, needed. Interestingly enough, given that the logic of quasi-truth is itself paraconsistent (see da Costa, Bueno, and French 1998), it's possible to develop a unified framework – the partial structures approach – that accommodates both (1) and (2). I briefly sketch this framework below.

The partial structures approach relies on three main notions: partial relation, partial structure, and quasi-truth (see, e.g., da Costa and French 1989, 1990, and 2003).[9] One of the main motivations for introducing this proposal comes from the need for supplying a formal framework in which the openness and incompleteness of information dealt with in scientific practice can be accommodated (see da Costa and French 2003). This is accomplished by two moves. First, the usual notion of structure is extended. In order to model the partialness of information we have about a certain domain, the notion of a *partial* structure is introduced. Second, the Tarskian characterization of the concept of truth for partial contexts is put forward, which leads to the corresponding concept of *quasi-truth* (or *partial truth*).

To introduce a partial structure, the first step is to formulate an appropriate notion of *partial relation*. When investigating a certain domain of knowledge Δ, we formulate a conceptual framework that helps us in systematizing and organizing the information we obtain about Δ. This domain is tentatively represented by a set D of objects and is studied by the examination of the relations holding among D's elements. However, we often face the situation in which, given a certain relation R defined over D, we do not know whether all the objects of D (or n-tuples thereof) are related by R. This is part of the

[9] Further developments and applications of the partial structures approach can also be found, e.g., in Bueno 1997, 1999, and 2000 and da Costa et al. 1998.

incompleteness of our information about Δ and is formally accommodated by the concept of partial relation. More formally, let D be a nonempty set; an n-place *partial relation* R over D is a triple $\langle R_1, R_2, R_3\rangle$, where R_1, R_2, and R_3 are mutually disjoint sets, with $R_1 \cup R_2 \cup R_3 = D^n$, and such that R_1 is the set of n-tuples that (we know that) belong to R, R_2 is the set of n-tuples that (we know that) do not belong to R, and R_3 is the set of n-tuples about which we do not know whether they belong or not to R. (Note that if R_3 is empty, R is a usual n-place relation that can be identified with R_1.)

However, in order to represent the information about the domain under consideration, we need a notion of *structure*. The following characterization, spelled out in terms of partial relations and based on the standard concept of structure, is meant to supply a notion that is broad enough to accommodate the partiality usually found in scientific practice. Partial relations do the main work, of course. A *partial structure* S is an ordered pair $\langle D, R_i\rangle_{i\in I}$, where D is a nonempty set, and $(R_i)_{i\in I}$ is a family of partial relations defined over D.

Two of the three basic notions of the partial structures approach have now been defined. To spell out the last and crucial one – quasi-truth – an auxiliary notion is required. The idea is to use, in the characterization of quasi-truth, the resources supplied by Tarski's definition of truth. However, since the latter is defined only for full structures, we have to introduce an intermediary notion of structure to "link" full to partial structures. And this is the first role of those structures that extend a partial structure A into a full, total structure (which are called A-normal structures). Their second role is purely model-theoretic, namely, to put forward an interpretation of a given language and, in terms of that interpretation, to characterize basic semantic notions. A-normal structures are defined as follows: Let $A = \langle D, R_i\rangle_{i\in I}$ be a partial structure. We say that the structure $B = \langle D', R'_i\rangle_{i\in I}$ is an *A-normal structure* if (1) $D = D'$, (2) every constant of the language in question is interpreted by the same object both in A and in B, and (3) R'_i extends the corresponding relation R_i (in the sense that each R'_i, supposed of arity n, is defined for all n-tuples of elements of D'). Note that although each R'_i is *defined* for all n-tuples over D', it is known to hold for some of them (the R'_{i1}-component of R'_i), and it's known not to hold for others (the R'_{i2}-component).

As a result, given a partial structure A, there are *too many* A-normal structures. We need then to provide constraints to restrict the acceptable extensions of A. In order to do that, a further auxiliary notion is introduced (see Mikenberg, da Costa, and Chuaqui 1986). A *pragmatic structure* is a partial structure to which a third component has been added: a set of accepted sentences P, which represents the accepted information about the structure's domain. (Depending on the interpretation of science that is adopted, different kinds of sentences are introduced in P: Realists will typically include laws and theories, whereas empiricists will add certain laws and observational statements about the domain in question.) A *pragmatic structure* is then a triple $A = \langle D, R_i, P\rangle_{i\in I}$, where D is a nonempty set, $(R_i)_{i\in I}$ is a family of partial relations defined over D, and P is a set of accepted sentences. The idea is that P introduces constraints on the ways that a partial structure can be extended (the sentences of P hold in the A-normal extensions of the partial structure A).

We can now formulate the concept of quasi-truth. A sentence α is *quasi-true* in A according to B if (1) $A = \langle D, R_i, P\rangle_{i\in I}$ is a pragmatic structure, (2) $B = \langle D', R'_i\rangle_{i\in I}$ is an A-normal structure, and (3) α is true in B (in the Tarskian sense). If α is not quasi-true in A according to B, we say that α is *quasi-false* (in A according to B). Moreover, we say that a sentence α is *quasi-true* if there is a pragmatic structure A and a corresponding A-normal structure B such that α is true in B (according to Tarski's account). Otherwise, α is *quasi-false*.

The idea, intuitively speaking, is that a quasi-true sentence α describes not the whole domain to which it refers, but only an aspect of it – the one modeled by the relevant partial structure A. After all, there are several different ways in which A can be extended to a full structure, and in some of these extensions α may not be true. As a result, the notion of quasi-truth is strictly weaker than truth: Although every true sentence is (trivially) quasi-true, a quasi-true sentence is not necessarily true (since it may be false in certain extensions of A).

To illustrate the use of these notions, let us consider a simple example. As is well known, Newtonian mechanics is appropriate to explain the behavior of bodies under certain conditions (say, bodies that, roughly speaking, have "low" velocity, are not subject to strong gravitational fields, etc.). But with the formulation of special relativity, we know that if these conditions are not satisfied, Newtonian mechanics is false. In this sense, these conditions specify a family of partial relations, which delimit the context in which the theory holds. Although Newtonian mechanics is not true (and we know under what conditions it is false), it is *quasi-true*; that is, it is true in a given context, determined by a pragmatic structure and a corresponding A-normal one (see da Costa and French 2003).

An important feature to note here is that a sentence *and* its negation can both be quasi-true. Of course, inconsistent sentences are not quasi-true in the *same* A-normal structure, but they can still both be quasi-true – as long as they are true in some A-normal structure. In other words, as defined above, if a theory is quasi-true, it is consistent (given that it is true in some full A-normal structure). But in some contexts, we may need to assert that an *inconsistent* theory is quasi-true. How can we do that?

Here is a way. If a theory T is inconsistent, we say that T is quasi-true in a partial structure A if there are "strong" subsets of T's theorems that are true in some A-normal structure. (We take "strong" to be a pragmatic notion, involving theories that are explanatory, have significant consequences, accommodate the relevant phenomena, etc.) In general, there are infinitely many "subtheories" of T that meet this condition. Of course, the interesting cases to consider are those in which A is a "good" pragmatic structure, in the sense that it reflects well the informal counterpart of T.

For example, let T be naïve set theory, formulated in first-order logic. In this case, in the pragmatic structure A, the set P of basic statements is constituted by statements that are typically taken to be unproblematic, such as the statement that asserts the existence of the union of two sets and the statement that expresses the comprehension schema restricted to a given set. In this case, there is only one relation in A, the membership relation, which is taken to hold for certain pairs of sets. Hence, T is quasi-true in A, given that there are several subtheories of T that are quasi-true, for instance, Zermelo-Fraenkel set theory, Quine's NF and ML, and von Neumann-Bernays-Gödel set theory.

Additional examples of this sort can be provided, for instance, with the earlier formulations of the calculus. Such formulations, articulated in terms of infinitesimals, were inconsistent, but again they have "strong" consistent subtheories. Similarly, even though the conjunction of quantum mechanics and relativity theory is inconsistent, it can still be quasi-true, given the existence of strong consistent subtheories.

Of course, this construction presupposes, for the usual reasons, a metatheory that is strong enough. Moreover, the construction is formulated in classical first-order logic, but it can be easily extended to higher-order logics, as in Frege's system, or to other logics, using the theory of valuations (see da Costa 1997).

Two points should be emphasized here: (1) The fact that inconsistent theories can be quasi-true *doesn't* entail that every sentence is quasi-true. After all, given a partial structure A, there exist sentences that aren't true in any A-normal structure. (2) The fact that inconsistent theories can be both quasi-true also doesn't mean that everything follows from the partial structures framework. After all, the logic of quasi-truth is paraconsistent (see da Costa et al. 1998). And as was pointed out above, in a paraconsistent setting, it's not the case that everything follows from an inconsistency. As a result, the partial structures approach provides the right sort of framework to examine issues regarding inconsistency in science. In terms of the approach, it's possible to represent, without triviality, inconsistent theories as being quasi-true.

Having said that, we can now return to the main issue under consideration, and discuss how the partial structures approach allows us to consider this issue in a new way.

4. INCONSISTENCY, BELIEF CHANGE, AND PRAGMATISM

The framework above indicates one way in which it's possible to make sense and pursue inconsistent belief systems without triviality. Such an outcome is far from idle, given the various benefits, discussed above, of taking seriously inconsistent belief systems. This outcome is also significant in a different way. Over the years, Levi has articulated a sophisticated and robust form of pragmatism (see, in particular, Levi 1991). And I think Levi's pragmatism can benefit from incorporating a more robust treatment of inconsistent belief systems. I turn to this topic now.

There are several components in Levi's pragmatism, and I cannot possibly do justice to all of them here. For our present purposes, though, I focus only on those components that bear on the issue of the aims of inquiry, since focusing on these components will be sufficient to illustrate my point.

According to Levi:

> The aim [of inquiry] is to find the true complete story of the world – that is, the complete story that is also error-free. Granting that the conception of a complete story is relative to a conceptual framework or a language used to represent potential states of full belief in that framework, how are we to understand truth or freedom from error *as an aim of inquiry*? I emphasize that we are concerned with truth as an aim of inquiry focused on revising doxastic commitments. (Levi 1991, p. 58)

Of course, as Levi insists, the true, error-free description of the world is achieved not in a single step, but basically through a process of *expansion* – where new information is added to a belief system (Levi 1991, pp. 71–116) – and *contraction* – where some information is excluded from the system (ibid., pp. 117–64). The crucial requirement of the proposal, however, is that *each stage* in the belief change process should *avoid error* (ibid., p. 161).

Given that error has to be avoided at each step, and given that contradictions are (in a classical setting) necessarily false, it's not surprising that Levi insists on the need for contracting as soon as we face a contradiction. As he points out, "expansion into inconsistency will . . . incur a maximum risk of error and, for this reason, will be resisted" (ibid., p. 90).

However, as Levi also emphasizes, to find the true complete story of the world, we need more than simply avoid error; we also need to search for *informative* descriptions of the world. But in searching for the latter, we often end up *incurring in error*. After all, the more informative a theory is, the less likely it is that the theory is true. This means that, to be able to obtain new and more information, we may have to risk error. As Levi puts the point:

> Even though inquirers should be concerned to avoid error, they also should be concerned to obtain new information of value, and such curiosity can justify risking error to obtain new information. (ibid., p. 160)

The idea of being justified to risk error in order to get new information is, of course, exactly right. Nevertheless, as the cases of Bohr and Frege mentioned above illustrate, sometimes the *only* way to obtain new information is by *entertaining inconsistent theories*. But Levi *resists* this move.

> What should be prohibited is being prepared to add new information to one's evolving doctrine when one is certain that it is false. No amount of new information can be worth the importation of certain error. However, this prohibition argues against expanding into contradiction by deliberate (= inferential = inductive) expansion; but it allows one to risk expanding into inconsistency via routine expansion, provided that the chances of importing contradiction are sufficiently low. (ibid.)

The point is clear. To deliberately import inconsistent information is never allowed; inconsistent information can be added to a belief system only via routine expansion, and even then, only if the risk of importing a contradiction is low enough.

Suppose, however, that the *only* way of obtaining certain bits of information is by *deliberately* expanding into inconsistency. Bohr's case discussed above illustrates this situation – as well as the earlier formulations of the calculus and the conjunction of quantum mechanics and relativity theory. To obtain the relevant information, in each of these cases, it looks as though one is *forced* to expand into inconsistency. Now, as long as the underlying logic is paraconsistent, there need not be anything unacceptable here. If we are searching for quasi-truth, these are all cases of inconsistent quasi-true theories. And by exploring the relevant partial structures, new information can be obtained, thus opening up the way for the formulation of better, consistent successor theories.

In other words, when faced with a contradiction, we need not embrace it, even in a paraconsistent setting. Given that we are *not* committed to the existence of *true* contradictions, eventually we will need to contract (or radically change) our inconsistent belief system. In this way, false information is *not* being *permanently* included in the system. The inconsistent information is included *provisionally* – the resulting system is taken only to be *quasi-true* – and all the information that depends on the inconsistency is tracked. And by exploring the resulting inconsistent but nontrivial system, we can make better informed decisions regarding exactly how, and when, to contract. In this way, the search for the true, error-free description of the world can be better implemented by allowing for a process of belief change that incorporates inconsistency.

5. CONCLUSION

Levi has developed an illuminating and systematic approach to belief change and the nature of inquiry. The considerations above suggest that, by making (more) room for inconsistency in his approach, he could achieve the pragmatist goals he has articulated so clearly in a more efficient way. Of course, this doesn't mean that he would need to adopt the whole framework outlined here (in terms of partial truth and partial structures), although this framework provides one way in which it's possible to combine inconsistency, informativeness, and avoidance of error in a systematic and unified form.

Perhaps all that is needed is just to adopt an underlying paraconsistent logic, given that this would allow one to explore inconsistent domains without triviality. As noted above, adopting such a logic *doesn't* amount to endorsing the existence of true contradictions, and so false information will not be permanently added to our belief system. All that is required is to use paraconsistent logic as a mechanism of consequence generation. In this sense, the logic is simply an engine of inquiry, a further tool in the exploration of the world. By incorporating such a tool, Levi's goals can be articulated still further, without any additional risk of error, and in a way that yields exactly the same results he obtains when we have a consistent domain (given that paraconsistent logic agrees with classical logic in consistent situations). In the end, this may not be a bad deal after all!

REFERENCES

Azzouni, J. 2004. *Deflating Existential Consequence: A Case for Nominalism*. New York: Oxford University Press.

Bohr, N. 1913. "On the Constitution of Atoms and Molecules." *Philosophical Magazine* 26: 1–25, 476–502, 857–75.

Boolos, G. 1998. *Logic, Logic, and Logic*. Cambridge, Mass.: Harvard University Press.

Bueno, O. 1997. "Empirical Adequacy: A Partial Structures Approach." *Studies in History and Philosophy of Science* 28: 585–610.

Bueno, O. 1999. “What Is Structural Empiricism? Scientific Change in an Empiricist Setting.” *Erkenntnis* 50: 59–85.
Bueno, O. 2000. “Empiricism, Mathematical Change and Scientific Change.” *Studies in History and Philosophy of Science* 31: 269–96.
Carnielli, W., and J. Marcos. 2002. “A Taxonomy of C-Systems.” In W. Carnielli, M. Coniglio, and I. D’Ottaviano (eds.), *Paraconsistency: The Logical Way to the Inconsistent*, pp.1–94. New York: Marcel Dekker.
da Costa, N. C. A. 1974. “On the Theory of Inconsistent Formal Systems.” *Notre Dame Journal of Formal Logic* 15: 497–510.
da Costa, N. C. A. 1997. *Logiques classiques et non classiques: Essai sur les fondements de la logique*. Paris: Masson.
da Costa, N. C. A., and O. Bueno. 1998. “Belief Change and Inconsistency.” *Logique et Analyse* 161–162–163: 31–56.
da Costa, N. C. A., and O. Bueno. 2001. “Paraconsistency: Towards a Tentative Interpretation.” *Theoria* 16: 119–45.
da Costa, N. C. A., and S. French. 1989. “Pragmatic Truth and the Logic of Induction.” *British Journal for the Philosophy of Science* 40: 333–56.
da Costa, N. C. A., and S. French. 1990. “The Model-Theoretic Approach in the Philosophy of Science.” *Philosophy of Science* 57: 248–65.
da Costa, N. C. A., and S. French. 2003. *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. New York: Oxford University Press.
da Costa, N. C. A., O. Bueno, and S. French. 1998. “The Logic of Pragmatic Truth.” *Journal of Philosophical Logic* 27: 603–20.
Feyerabend, P. K. 1988. *Against Method*, revised ed. London: Verso.
Frege, G. 1950. *The Foundations of Arithmetic*. English translation by J. L. Austin. Oxford: Blackwell. Originally published in 1884.
Frege, G. 1962. *Grundgesetze der Arithmetik*, 2 vols. Hildesheim: Georg Olms. Originally published in 1893 and 1903.
Gärdenfors, P. 1988. *Knowledge in Flux*. Cambridge, Mass.: MIT Press.
Hale, B., and C. Wright. 2001. *The Reason’s Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Clarendon.
Lakatos, I. 1978. *The Methodology of Scientific Research Programmes*, ed. John Worrall and Gregory Currie. Cambridge: Cambridge University Press.
Levi, I. 1991. *The Fixation of Belief and Its Undoing*. Cambridge: Cambridge University Press.
Levi, I. 1996. *For the Sake of the Argument*. Cambridge: Cambridge University Press.
Levi, I. 2003. “Contracting from Epistemic Hell is Routine.” *Synthese* 135: 141–64.
Mikenberg, I., N. C. A. da Costa, and R. Chuaqui. 1986. “Pragmatic Truth and Approximation to Truth.” *Journal of Symbolic Logic* 51: 201–21.
Olsson, E. 2003. “Avoiding Epistemic Hell: Levi on Pragmatism and Inconsistency.” *Synthese* 135: 119–40.
Priest, G. 1987. *In Contradiction*. Dordrecht: Nijhoff.
Priest, G., R. Routley, and J. Norman, eds. 1989. *Paraconsistent Logic* Munich: Philosophia Verlag.
Wright, C. 1983. *Frege’s Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.

6

Levi on Risk

Nils-Eric Sahlin

In Plato's *Apology*, Socrates declares:

> Probably neither of us knows anything really worth knowing: but whereas this man imagines he knows, without really knowing, I, knowing nothing, do not even suppose I know. On this one point, at any rate, I appear to be a little wiser than he, because I do not even think I know things about which I know nothing.

First-order knowledge is important, but second-order knowledge of what one does or does not know is even more important: That is, it is essential to have what we might call "epistemic self-knowledge." Scientists are knowledge-driven. That is why inductive methods are so popular in science. Knowledge is a good thing, but there are situations in which we require more – in which we require wisdom and thus epistemic self-knowledge. For example, when our theories fail to deliver results, when we are forced to find new theories or create new hypotheses, it is vital to know what one does not know.

Indirectly, Socrates tells us something about rational decision making. When everything is propitious, when we can represent our knowledge and our values with unique probability distributions and precise utility functions, we can simply maximize expected utility. We can use one of the classical theories – for example, Ramsey's, or Savage's, or Jeffrey's theory. But when we are uncertain about the extent of our knowledge, when things are indeterminate, or when we are uncertain about our preferences, we know that the traditional theories will be of little or no use to us.

Isaac Levi saw this early, before others, and developed an alternative to the classical theories of rational decision making. Levi's theory allows the agent to be uncertain (or indeterminate) in his probability assessments, and to have equivocal preferences. In this chapter I do not discuss the pros and cons of Levi's decision theory. I have done that elsewhere. Instead, I want to focus on one of his less well-known papers: "A Brief Sermon on Assessing

Accident Risks in U.S. Commercial Nuclear Power Plants." This piece was written in the late 1970s and published as an appendix in *The Enterprise of Knowledge*. In it, Levi indicates that his theory can be used to improve risk analysis and risk management. However, without a basic understanding of Levi's decision-theoretical approach, it is difficult to appreciate his insights fully.

LEVI'S DECISION THEORY

A decision maker's information at a certain time about states of nature is, in Levi's theory, represented by a convex set of probability distributions (denoted *B*). Distributions in the set are the "permissible" distributions. Traditional theory of rational decision presupposes that *B* contains one and only one distribution. It is assumed that a decision maker's beliefs about states of the world in a given situation can be represented by a unique function defined over those states. But it can be argued that this misrepresents the decision maker's knowledge and beliefs and that the quality and quantity of the decision maker's information is not fully accounted for. Levi's approach provides a far more adequate and complete representation.

Levi also generalizes traditional theory by introducing a set of "permissible" utility functions (denoted *G*). Traditional theory assumes that there is but one utility function representing the decision maker's preferences (not counting affine transformations). A set of utility functions tells us far more about the uncertainty and robustness of the decision maker's preferences than a single function does.

The traditional decision rule says:

> In a given decision situation, choose the alternative with maximal expected utility, i.e. maximize expected utility.

This apophthegm does its work as long as the decision maker's beliefs and values can be represented by a unique probability distribution and a single utility function. But with sets of measure the situation is different. You cannot maximize over intervals or sets. Generalized decision theories therefore have to offer new decision rules. Levi suggests we use a three-step procedure, a lexicographically ordered set of rules.

First, a decision alternative is said to be *E-admissible* if and only if there is some probability distribution in the set *B* (which Levi assumes to be convex) and some utility function in the set of utility functions *G* such that the expected utility of the action alternative relative to the two distributions is maximal among all the available action alternatives. It is then claimed that a decision alternative must be E-admissible in order to be choiceworthy.

Second, an alternative is *P-admissible* if it is E-admissible and it is "best" with respect to E-admissible option preservation among all E-admissible options. This condition has to do with the possibility of deferring decision. "[T]he injunction to keep one's options open is a criterion of choice that is based not on appraisals of expected utility but on the 'option-preserving' features of options." It is worth observing that this rule may contract as well as expand the set B. Without due care, it may even turn B into a nonconvex set.

Third, and finally, a P-admissible alternative is *security optimal* if and only if the minimum utility value assigned to some possible outcome of the decision alternative is at least as great as the minimal utility value assigned to any other P-admissible alternative. And a decision alternative is *S-admissible* if it is E-admissible, P-admissible, and security optimal (relative to some of the decision maker's utility functions). It is the S-admissible alternatives that are admissible for final choice. If there is more than one admissible alternative, a die, or a similar tool, has to be used.

After Ramsey's work in "Truth and Probability" (of which von Neumann and Morgenstern's utility theory and Savage's decision theory can be counted rediscoveries), Levi's theory makes one of the few profound and substantial contributions to probability and decision theory: It is an indisputable breakthrough. One of the theory's great attractions is, of course, the way it represents our beliefs and desires. Another admirable feature is the carefully developed set of decision rules; and the way in which Levi's theory solves important paradoxes due to Allais and Ellsberg is a genuine advance.

But like all good theories, Levi's is not without blemishes. The assumption that B is convex is counterintuitive and sometimes problematic. The theory is not stable under the fusion of states; and in it, Luce and Raiffa's classical Axiom 7 is violated. Generalized theories must eschew either independence axioms or ordering axioms, and the choice here is less straightforward than one might think (see Seidenfeld's well-known paper). Again, there are rival theories, also aiming at a more complete representation of beliefs and utilities, taking indeterminacy or unreliability into account; and it can be shown that these competitors do not always give the same recommendation as Levi's theory. Compare, for example, the theories of Levi, Kyburg, and Gärdenfors and Sahlin. And if we cannot refute two of these, which are we to use as a basis for action? But these problems can hardly be seen as serious blemishes; they are more like charming freckles.

With Levi's decision theory before one, it is rather easy to see that there are four paradigmatic types of decision. For mnemonic advantage, I use the expressions "Type 1," "Type 2," "Type 3," and "Type 4" in connection with situations, decisions, and decision situations.

In *Type 1* situations, the agent has extensive knowledge and information. He also knows what he likes and dislikes (i.e., has clear and distinct preferences). This is represented by a unique probability distribution and a single utility function.

In *Type 2* situations, the quality and quantity of information is weak (indeterminacy), but the decision problem is one in connection with which the agent still has clear and distinct preferences. This is represented by a (convex) set of probability distributions and a single utility function.

In *Type 3* situations, the quality and quantity of information is good, but the agent lacks clear and distinct preferences. This is represented by a unique probability distribution and a set of utility functions.

In *Type 4* situations, both information and preferences are indeterminate or unreliable. This is represented by a (convex) set of probability distributions and a set of utility functions.

Type 1 decision problems are the paradigm cases with which classical theories of rational choice and decision-making deal. Theories such as Ramsey's, Savage's, Jeffrey's, and von Neumann and Morgenstern's are all developed to deal with decision situations of this kind. But the traditional theories are ill-equipped to handle the three other types of situation. We all know that there is more to decision making than asking oneself whether to break or not to break an egg into a bowl, or whether to take left or right at a cross-roads to get to the pub, and pretending that we "can determine to any degree of accuracy whatsoever, the probability . . . that" the egg is rotten (Savage), or that we can work out "to seven places of decimals a result only valid to two" (Ramsey) – all for the sake of maximizing expected utility.

The hard choices are made in Type 2, 3, and 4 situations. In these situations the traditional theories are not great guides to action, so we need Levi's theory or something similar. We need to be able to represent indeterminate beliefs and imprecise values. We need to work with sets of measures rather than single functions. And we must allow for the introduction of a more complex but also more complete decision procedure.

Risk is a vague concept. There is not one but literarily hundreds of more or less well-thought-out definitions. The general view is that risk has something to do with the outcomes of our actions – with their character, seriousness, and likelihood. Thus risk can be taken to mean something like "an outcome below an aspiration point," "the probability of a negative outcome," "the frequency with which a negative outcome occurs," "negative utility," or information about "objective" probabilities, or simply seen as an epiphenomenon of the shape of the utility function (of its convexity).

But there is more to risk than outcome risk. And this is what Levi's paper is all about. To ground a decision on scanty and indeterminate information and/or unclear and jumping preferences, is to take a risk, but not an outcome risk. What we take is an *epistemic risk*. We are not gambling with outcomes but with our beliefs and values.

Numerous cases (e.g., nuclear power plants, mad cow disease, genetically modified food, electromagnetic fields) have taught us that external factors can be involved in the creation of epistemic risk. Reliable risk analysis requires careful scrutiny of, so to speak, the present epistemic state. It is not enough to identify and evaluate outcome risks; an estimation and evaluation of the prevailing state of ignorance (indeterminacy) is also needed. One factor known to produce epistemic risk is "the unreliable research process." Research is a dynamic process that, for various reasons, occasionally gives us incorrect or indeterminate results. The quality of its results depends on the standard of the scientific machinery involved. To look solely at the results, and not ask whether the machinery has worked properly, is to leave epistemic risks out of consideration.

What we might call "the cyclopean perspective" is another factor in epistemic risk. Our search for knowledge can, and sometimes does, make us one-eyed. Psychologists, neurologists, and philosophers have taught us that we tend to look for evidence that confirms what we believe – that we are reluctant to look for falsifying evidence. Francis Bacon was, for example, well aware of this difficulty. But this one-eyed outlook can produce, and has produced, serious epistemic risks.

A third factor in epistemic risk is "the unrealizable research process." We might get caught in situations where it is, for moral or practical reasons, difficult to carry out controlled experimental studies. As a result we might have to rely on indirect evidence rather than solid and direct empirical evidence.

Consider this illustration. The best way to find out the effects of toxic substances on humans is to conduct experiments on humans, not animals. But that is often impossible, because the tests would be immoral or in some other way unacceptable. Therefore we have to rely on animal experimentation (which is far from unproblematic from a moral point of view), and hence on indirect evidence. Our moral commitments thus create uncertainty. They produce epistemic risk. But there are also practical problems. In testing toxic substances, one works with three dose groups: a control group (zero dose) and two groups of animals exposed to higher or lower doses of the substance. In a twenty-eight-day dose toxicity test, five animals of each sex and dose group are used, that is, thirty animals in total. To get significant results with groups of that size, the experiment must be carried out at high dose levels, and these levels are well above those we are normally exposed to. In the light of this type of experiment, it is difficult to say what increase in risk can be expected at the normal dose level. In addition to statistical limitations and the limitations of the experimental design, there is the problem of extrapolation. The hamster is, for example, 10,000 times less sensitive to dioxin than the guinea pig, and man is neither a hamster nor a guinea pig. Indeterminacy is the result.

In "A Brief Sermon on Assessing Accident Risks in U.S. Commercial Nuclear Power Plants," Levi does not primarily focus on external epistemic risk-producing factors. Instead, he points to a set of internal factors. An advocate of the traditional type of decision theory will be blind to these problems, because within such a theory some of the most serious problems are idealized away. But in the framework of a more complete theory such as Levi's, the epistemic difficulties can be spotted.

Structuring a decision problem can be difficult. We must identify the possible acts, the possible states of nature, and the relevant outcomes. When we draw the decision tree, we want it to be, not only complete, but set out in detail. Two decision trees can, of course, represent the same decision problem, even if one has, for example, fewer states of nature than the other. In practice, how the tree is drawn, how the problem is structured, depends on what we know about the world. That knowledge will influence our assessment of probabilities and utilities. It will influence the epistemic risks we take and even put calculated decision making at risk. In Levi's words: "In any case, failure to account for all serious and relevant possibilities threatens all deliberate decision-making. This does not mean that we should not attempt to be as clear as we can about such possibilities. It means only that we should avoid congratulating ourselves too much and looking at our analyses as security blankets."

Levi also allows us to see that, being driven by a desire to maximize (because that is what our theories in one way or another ask us to do), we cheat ourselves into a kind of faith in robustness – into believing that there is a unique probability distribution and single utility function. In point of fact, however, we face a good deal of inconspicuous unreliability.

Statistics and traditional methods of decision analysis force us to pretend that we are in Type 1 situations when we in fact are stuck in the opposite corner dealing with Type 4 decisions. But we need to "look for evidence supporting the elimination of one or the other of [the] distributions," acknowledge the lack of robustness and unreliability, and employ a more complete type of decision theory.

Type 4 situations can involve very bad outcomes with very low probabilities – so low indeed that it is difficult to conduct experiments that will give us precise information about the "true" probability distribution. It is, for example, impossible in principle to conduct controlled experiments that disclose the effects of very low doses of radiation. Weinberg writes that "to determine at the 95 percent confidence level by a direct experiment whether a 150 millirems will increase the mutation rate by 0.5% requires about 8 000 000 000 mice." In situations like this, where we cannot conduct the necessary experiments and have to rely on indirect evidence, more than one rival hypothesis (distribution) will fit the available data. If we are obliged by "formal," "aesthetic," or "pragmatic" considerations to maximize and choose just one of the distributions, we will take unnecessary epistemic risks.

Addressing this problem, Levi summarizes: "The moral of the story is that we should learn to suspend judgment. We should . . . learn to acknowledge that the data justifies and, indeed, obligates us to suspend judgment concerning the objective chance distribution. . . . We should be prepared to adopt creedal states of hypotheses . . . which are indeterminate and which allow many diverse distributions to be permissible."

Risk communication is, as they say, a hot topic. Why do members of the public often refuse to accept expert risk assessments? What causes them to question the specialist's knowledge and values? The conflict between the public and the so-called experts is multidimensional. Economic, psychological, sociological, and educational reasons lie behind it. In his article, Levi uses his decision theory to help us to understand an important aspect of the problem. If scientists, owing to the limitations of the tools and methods they use, are forced into treating serious decision problems as if they are always of Type 1, and if the public, sensitive to the overtones of indeterminacy and unreliability that risk assessments carry with them, look on the same problems as Type 4 problems, then we have a potential conflict. Well-known examples are the nuclear safety debate and the BSE and GMO controversies. And you do not transform a Type 4 problem into a Type 1 problem by eating hamburgers containing British meat. To avoid conflicts of this sort it is necessary to take a Socratic approach to decision making – to admit that there are epistemic uncertainties, and not to pretend that the set of permissible distributions is narrower than it really is.

The experts have to gain trust by giving a clear account of all sorts of indeterminacy. When it comes to the set of permissible probability distributions, research can help us to rule out single distributions and even sets of distributions. Or, alternatively, that research might encourage us to expand what careful deliberation has shown to be too narrow a set of permissible distributions. When it comes to probability distributions, our aim is to track truth. We are gambling with truth, and that, at least in theory, makes information swapping "easier." However, when we are talking about sets of utility functions the situation is far more difficult. Someone in a Type 1 situation who firmly believes that the single utility function u(.) correctly represents his preferences and desires cannot be convinced by "facts" to change his or her mind. We have left the business of truth tracking. What Levi's theory gives us is, first, an understanding of why communication conflicts emerge, and, second, tools and insights that can help to mitigate and prevent disagreements. Effective and honest risk communication presupposes a Socratic approach to decision making within which the experts admit and communicate indeterminacy.

There is yet another important lesson to be drawn from Levi's theory. The precautionary principle says: "In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation" (Article 15 of the Rio declaration of 1992). Today the precautionary principle is used as a guideline for environmental decision making.

It has been argued that the principle is vague and not really applicable as it stands, that it is a principle in need of interpretation. The precautionary principle cannot be straightforwardly compared with other decision principles. It is a disconnected principle, a free-floating rule or guide to conduct. Such principles, whether they are decision axioms or moral maxims, tend to conflict with other reasonable principles. A solid theoretical framework is lacking.

The principle has two faces: a value side and a belief side. Note, however, that the principle is in another sense one-sided: It focuses on negative outcomes, negative values. And other formulations of the principle also emphasize negative consequences such as "threats" and "harm": "When an activity raises threats of harm to the environment or human health, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically" (Wingspread 1998). On the other hand, the belief part of the principle is interesting since it emphasizes indeterminacy, ignorance, and unreliability. And this is what connects it with Levi's theory and ideas.

The precautionary principle grew out of dissatisfaction with traditional methods of risk analysis (decision analysis and cost-benefit analysis). The type of risk analysis and risk management methods that Levi scrutinizes in his paper – methods putting too much weight on maximization and none on indeterminacy – can hardly be counted as precautionary. Levi's theory, however, is built around several precautionary ideas.

That "some cause and effect relationships are not fully established" is in Levi's theory handled by the set of permissible probability distributions. If we are uncertain about causes and effects, then the set of permissible distribution will contain a wider range of distributions than it does when we have a clear picture of the causal network. And the theory does not force us to squeeze the alternative distributions into a single distribution (and so does not force us to take maximal epistemic risks) by neglecting all but one of them or by trying to construct a "composite" distribution using an ad hoc method weighting the first-order distributions. The indeterminacy aspect of the precautionary principle is, in fact, nicely handled by Levi's theory.

But Levi's theory introduces precautionary measures at different levels. His first decision rule, E-admissibility, is, of course, not what we think of as a precautionary principle. But his second decision rule, P-admissibility, which emphasizes the possibility of deferring decision and urges that one ought to keep one's options open, is a rule with a strong precautionary flavor. Equally, security optimality puts the emphasis on negative outcomes: "harms," "threats," "irreversibility," and so on. Thus it embodies that part of the precautionary principle.

The theory also points to another precautionary problem. When we are facing a hard choice, it is not only our beliefs that may be indeterminate. Our preferences and values are often vague and undetermined as well. Do we really need or want the new technology? Do I like the brave new world that science offers? A precautionary principle should look not only at the indeterminacy of our beliefs and knowledge, but also at the instability of our preferences. The precautionary principle does not do this, but Levi's theory does.

Levi's theory appears, then, to give us the tools and techniques required for sound risk analysis and effective risk management. It shows that we do not have to be satisfied with anything less than a complete decision theory. I began this article by citing Socrates, and I will close it by quoting Levi:

> But the demands of such inquiry require that experts admit the limits to what they know. Scientists and technologists should not pretend to a knowledge they do not have because a government or a public demands that they be supplied with answers

to questions for which there is insufficient evidence. And the public and government should understand and respect the limits on what they can expect of responsible scientists and engineers. . . . Above all we should beware of epistemologies which permit us to violate this counsel and indulge our tastes for the familiar, simplicity, explanatory power, naturalness, paradigmatic methods of puzzle solving, and the like without regard to the risks of error both in theory and in practice which our indulgences may be incurring.

REFERENCES

Gärdenfors, P., and N.-E. Sahlin. 1988. *Decision, Probability, and Utility.* Cambridge: Cambridge University Press.

Kyburg, H. 1983. "Rational Belief." *Behavioral and Brain Sciences* 6: 231–73.

Levi, I. 1980a. "A Brief Sermon on Assessing Accident Risks in U.S. Commercial Nuclear Power Plants." In *Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants: Appendix 1, Accident Definition and Use of Event Trees, and Appendix 2, Fault Trees*. Washington, D.C.: Nuclear Regulatory Commission, October 1975. Republished in Levi 1980b, pp. 431–44.

Levi, I. 1980b. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. Cambridge, Mass.: MIT Press.

Luce, R. D., and H. Raiffa. 1957. *Games and Decisions*. New York: John Wiley & Sons.

Sahlin, N.-E., and J. Persson. 1994. "Epistemic Risk – The Significance of Knowing What One Does Not Know." In B. Brehmer and N.-E. Sahlin (eds.), *Futures Risk and Risk Management*, pp. 37–62. Dordrecht: Kluwer Academic.

Seidenfeld, T. 1988. "Decision Theory without Independence or without Ordering, What Is the Difference?" *Economics and Philosophy* 4: 267–315.

Weinberg, A. M. 1972. "Science and Transscience." *Minerva* 10, no. 2: 208–22.

"Wingspread Statement of the Precautionary Principle." 1998. Http://www.safe2use.com/ca-ipm/01-03–30.htm.

7

Vexed Convexity

Henry E. Kyburg, Jr.

1. THE HORSE OR THE CART?

John Maynard Keynes (Keynes 1952) proposed that probability should be *legislative* for rational belief. He also proposed that probabilities should form only a *partial* order: There were to be incomparable pairs of probabilities where the first is not larger than the second, the second not larger than the first, yet the two probabilities are not equal.

Frank Plumpton Ramsey objected (Ramsey 1931), quite correctly, that any such scheme depended on being able to *relate* beliefs and probabilities. He disregarded the second proposal, and so took probabilities to be numbers, so that what he took to be necessary was a way of *measuring* degrees of belief.

Ramsey offered a somewhat naïve operational way of measuring beliefs. He himself took it to be no more than approximate ("I have not worked out the mathematical logic of this in detail, because this would, I think, be rather like working out to seven places of decimals a result only valid to two" (ibid., p. 180). What was important about Ramsey's proposal was that it also suggested *why* beliefs (assuming their measurability) should satisfy the probability calculus.

Ramsey's approach became the model for later "subjectivistic" approaches. First, we think about ways in which to measure degrees of belief; second, we consider why those degrees should satisfy the probability calculus; and third, we consider how those probabilities should be updated in the light of new evidence. Each of these features has been changed or generalized, but it is nearly universal that Keynes's arrangement of cart and horse has been replaced by Ramsey's: The cart (the measurement of degrees of belief) has been placed first, and the horse (the issue of rationality) has been placed second.

By this I don't mean merely that issues of rationality are regarded as less important; I don't think that is the case. However important those issues are,

This material is based on work supported by the National Science Foundation under Grant IIS 0082928.

it is thought that the material with which we have to work consists, from the outset, of degrees, or perhaps intervals, of belief.

Keynes's view was that, on the contrary, we begin with a *logical* relation between pairs of sentences, discover (in some not very clear way) how different instances of this relation are related, and deliberately achieve rationality by ensuring that our beliefs *conform* to those relations. This view, of course, presupposess that the probability relation is a matter of logic, and therefore presupposes a particular view of logic. It presupposes that our partial beliefs (at least) are subject to our wills: for example, that no matter how strong a "hunch" we have that the coin will land heads, we will have a belief of about a half in the eventuality of heads. Some people deny this (van Fraassen 1984). Many people, perhaps Ramsey among them, think logic should be restricted to matters of consistency.

Levi is one of those who has stretched Ramsey's view in a number of directions (Levi 1967). Rather than taking Ramsey's measurement procedure too seriously, he allows that degrees of belief can be approximate, *imprecise*. The decision theory in which these imprecise degrees of belief, or imprecise probabilities, play a role is complicated and controversial, but certainly richer than that anticipated by Ramsey. And updating may occur by temporal conditioning, but may also occur in other ways. Nevertheless, the agent's set of beliefs, or set of dispositions to act, comes first and is examined for conformity to certain standards before being given the honorific "rational." Among the constraints on rationality that Levi imposes [Levi 1999, p. 512] are two that concern us. "Credal Coherence" is the doctrine that the credal state of an agent can be represented by a *set* of conditional probability functions $Q(.|x)$ for every x consistent with the evidence and background knowledge K. "Credal Convexity" is the condition that for each x consistent with K, the set of permissible Q functions is convex that is, contains any linear combination of Q functions that it contains. [ibid., p. 512].

2. THE HORSE OF RATIONALITY

Let us suppose that the horse of rational constraints on what we might call propositional commitment comes first, and that beliefs follow in the train of these constraints. What constraints might these be?

2.1. Probability

The constraints that concern us are those bearing on what has been called partial belief. I shall assume that there are objective grounds for determining the probability of an arbitrary statement S in a formal language, relative to a set of statements K accepted as evidence in that language. Roughly speaking, the grounds are like these:

1. S is known (in K) to be equivalent to a statment of the form $\ulcorner a \in T \urcorner$ – that is, there are terms a and T in the language such that $\ulcorner S \equiv a \in T \urcorner$ is in K.
2. There is a term R in the language such that the sentence $\ulcorner a \in R \urcorner$ is in K. We call the extension of R a reference class for S.
3. There is approximate statistical knowledge in K connecting R and T. We represent this by a statement of the form $\ulcorner \% \overline{x}(T(\overline{x}), R(\overline{x}), p, q) \urcorner$, which is to be read: The proportion of tuples satisfying T among those satisfying R lies between p and q.

Now for a given sentence S there may be several collections of terms a, R, and T that satisfy conditions (1)–(3). Elsewhere we have given procedures for elminimating from consideration some of these triples $\langle a, R, T \rangle$. But there may still be some left. We take the *probability* of S to be the cover of the uneliminated intervals.

One could worry about whether this interval, in real life, would be so broad as to be unhelpful. In deciding whether or not to act on S, it isn't much help to know that its probability is [0.1,0.9] (but perhaps it is better than nothing). To get more insight into this question, we need to calculate probabilities relative to quite rich bodies of knowledge. For the moment, to get on with the argument, we simply assume that it is possible to arrive at reasonable probabilities along the lines suggested in Kyburg and Teng (2001).

What is most important about these rational constraints is that they are objective. They are objective in two ways. First, they are logically objective in the sense that given a body of evidential knowledge, the probability of every sentence in our language is determined. Thus, if two people share the same body of evidence, they will assign the same probabilities to sentences. Second, probabilities are empirically objective in the sense that every probability is based on *known statistical facts*.

Logical objectivity is what renders probability socially useful in a climate where people are capable of sharing, and willing to share, information and data. Of course, in order for two people to agree on a probability for a statement S, they need not share *all* their evidence; they have to share only what bears on the statement in question, and the enumeration above provides an explicit characterization of "bearing on S."

Empirical objectivity is even more important than logical objectivity. It is the foundation in empirical relative frequencies that provides a link between probabilities and the world and that constitutes the main reason that one should base one's choices on mathematical expectation, where one can. It can be shown that if the probability of S is $[p, q]$, then the proportion of possible future worlds compatible with one's body of evidence in which S is true lies between p and q. If probabilties are very imprecise, then of course expectations will be also, and thus may not provide much guidance. We return to expectation below.

All this is relative, of course, to a canonical first-order language. Given the language, probabilities are determined by evidence. But one of the ways in which science progresses is by changes in language (in with "oxygen," out with "phlogiston"). We do not have a full account of rationality until we have an account of the grounds for changing languages. That is not any part of this present project, though some suggestions were thrown out in Kyburg (1990).

To return to the horse. The horse of rationality that powers belief, decision, and deliberate behavior is probability. Probability is both logically objective in the sense that given a body of evidence, there is one and only one probability for a given sentence S, and it is empirically objective, in the sense that the probability of a sentence S is $[p, q]$ relative to a body of knowledge only if the proportion of worlds compatible with that body of knowledge in which S is true lies between p and q.

2.2. Conditional Probability

In one sense, as Keynes remarked (1952), all probabilities are conditional – they are conditional on what we know. The updated probability of S in the light of new evidence represented by T is thus just the probability of S relative to the body of evidence *K augmented by T*. Exactly how augmentation is to work is another question. This is not the sense of conditioning that most people are interested in. What they want to know is when we can update a probability by dividing one old probability by another old probability. When probabilities are construed as *intervals* it is not clear that this operation can be given a useful sense.

There are cases, however, in which the probability of S relative to $K \cup \{T\}$ is close to the quotient of the probability of $S \wedge T$ relative to K, and the probability of T relative to K. That is, the probability interval of $S \wedge T$ is very close to the number r, and the probability of T is very close to the number t, and the probability of S relative to K augmented by T is close to r/t.

This holds only if a number of conditions hold, which we need not go into here. One frequent condition that is satisfied when conditioning in this sense holds is worth noting, however. If S has the form $\ulcorner a \in T_s \urcorner$ and T has the form $\ulcorner a \in T_t \urcorner$ and T, and both probabilities have a common reference class R, then we may make use of the mathematical fact that the frequency of T_s in $\ulcorner R \cap T_t \urcorner$ is always the ratio of the frequency of $\ulcorner T_s \cap T_t \urcorner$ in R to the frequency of T_t in R, provided, of course, that the latter frequency is not 0.

Note that this does *not* mean that when we augment the sentences K by adding the sentence $\ulcorner a \in T_t \urcorner$, the probability of $\ulcorner a \in T_s \urcorner$ is constrained to be close to the ratio of a number in the probability interval appropriate to $\ulcorner a \in T_s \cap T_t \urcorner$ to a number in the probability interval appropriate to $\ulcorner a \in T_t \urcorner$. Perhaps we have reason to believe that if the first roll yields a three, that is a sure sign that the die is biased.

In any case, something approaching classical conditioning depends on a number of assumptions that are not always satisfied. These assumptions *may* be satisfied in games of chance. But while they reflect foundational relations among frequencies, they do not, in the first instance, represent foundational relations among probabilities.

3. THE CART OF BELIEF

We have so far not even hinted that there is a connection between probability and belief. There is, of course. Rational belief is belief that in some sense conforms to probability or that in some sense is constrained by probability. But as Ramsey (1931) noted, in order to make one thing conform to another, the two things must share a common currency. Ramsey's solution was to gloss over the fact that Keynes did not take probabilities to be completely ordered, so that he could suppose that probabilities were real numbers, and to give a behavioristic interpretation of belief, so that "degrees of belief" could also be measured by real numbers. God bless the number system. Of course if you identify probabilities with real numbers, then Keynes's vision (1952) is no longer very compelling, and all that survived Ramsey's analysis was that degrees of belief, as measured by real numbers, should obey the *relations* embodied in the probability calculus.

I have argued that there are serious problems with the attempt to produce a behavioristic interpretation of belief (Kyburg 2003). If we force an agent to post odds and take bets, this "forcing" must be taken account of. If we attempt to infer an agent's probabilities from his or her choices, we are taking as revelatory choices the agent him- or herself takes to be arbitrary.

Belief comes in two forms: There is *categorical* belief, generally expressed in categorical statements, such as "There is a crow on the barn," "The early flight to Dallas is usually late," "Dogs are carnivorous," "Margaret is in Syracuse," "In the long run, this die would land with one up about a sixth of the time." These are the kinds of things one might claim to believe; one would not, in ordinary circumstances, think about betting against them. (Being offered odds of a million to one is not an ordinary circumstance – and is probably not itself credible.) Then there is *partial* belief, expressed in a wide variety of ways, such as "I'm almost sure it will rain before the day is out," "There is probably a newborn calf this morning," "I seriously doubt that James will show up," "Charles may or may not come to work" (as a categorical asssertion this would be a tautology), "Eight'll get you ten that the sun will come out before noon," "You are almost certain to lose money at the casino tonight."

I'm displaying examples not because I am particularly concerned with the "ordinary language" meanings of "belief" but simply to suggest that there are two generally different usages that are both useful in our intellectual economy. We take the categorical usage to represent the acceptance of statements of our language as *evidence*; the statments in the body of evidence K are statements that are accorded "full belief." This does not mean that such statements are incorrigible; further evidence might lead to their excision from K. As the idea of a million to one bet shows, the line between "full" and "partial" belief is not hard and fast. And of course it makes perfectly good sense, in some contexts, to ask for the *evidence* that supports an evidential certainty in K. Despite all this, we shall go ahead and presuppose the following epistemic structure, since it provides a handy framework for our later discussion:

1. In a given context there is a set of fully believed evidential certainties, K.
2. In the same context, there is a set of statements to which less than full belief is assigned.

We have no reason whatsoever to think that beliefs of the second sort can be characterized by real-valued degrees. On the contrary, introspection (for what it's worth) suggests that while there are sometimes relations of "greater" or "less" between beliefs (I think it more likely that it will rain tomorrow than that it will snow), it is not the case that beliefs can be ordered. (There is no real number k such that I think it exactly k times as likely to rain as to snow tomorrow.)

This is not to say that we could not have a theory of behavior that made use of such abstract entities as degrees of belief. I don't think we have one, though I suppose there are people who would disagree with me. Bayesian decision theory is not a theory about how people *do* behave, nor is it a theory of how people "ought" to behave unless we have a way of specifying what degrees of belief they ought to have. In any event, though, real valued degrees of belief should not be taken as a starting point. We should take these "degrees" of belief as somewhat amorphous objects that admit, at best, only a partial order.

(Note that comparative degrees of belief do not address the issues we are considering. Comparative degrees of belief admit of comparison, as do amorphous degrees of belief, but the temptation to suppose something like "if A is not bigger than B, and B is not bigger than A, then A and B must be the same" is almost overwhelming, and then we have essentially a complete order.)

Since "degrees" carries with it a presumption of real number measurability, we henceforth speak of *lumps* of belief. Some lumps, clearly, are bigger than others; some pairs of lumps are just incomparable – one cannot say that one is larger than the other, nor yet that the other is larger than the one.

4. CONVEXITY

4.1. Convexity in Frequency

Statistics are not always convex: We may know that either one of three of the tosses of this coin land heads or that two of three of the tosses of this coin land heads. If that is what we know, we know that the frequency of heads is not one-half. The *probability* of heads (as we have defined it) is the same interval, $[\frac{1}{3}, \frac{2}{3}]$, as is the probability of heads on the toss of a coin whose bias leads it to land heads anywhere from one-third to two-thirds of the time.

Put more formally, what we know of a distribution of a quantity in a reference class is often, but not always, that it is one distribution in a set of distributions that is convex in one or more parameters. What we know of the coin may be that the distribution of heads in sets of its tosses is binomial with a parameter p between $\frac{1}{3}$ and $\frac{2}{3}$. Or what we know may be that it is either binomial with parameter $\frac{1}{3}$or binomial with parameter $\frac{2}{3}$.

It is possible for a set of distributions to be convex in one parameter and not in another, or in some collection of parameters. We might know that the distribution of the weights of fish in a lake is normal with a mean between μ_1 and μ_2 and a fixed standard deviation σ, or we may know only that σ lies in a certain interval. Of course we may also know that the distribution is one of two (or k) distinct distributions. For example, we may know that all of the fish come from one of k hatcheries, each of which produces fish with a characteristic distribution of weights.

It is ultimately our knowledge of distributions on which our (interval) probabilities depend. This knowledge is generally of sets of distributions, and these sets of distributions are often convex in one or more parameters, particularly when they embody knowledge inferred directly from statistical data. This, I think, is where the idea of convexity comes from, but of course it has no bearing on belief – even rational belief – except indirectly through the probabilities involved.

4.2. Convexity in Probability

Given that probabilities are intervals, what would it mean to say that probabilities are convex? We'd have to be mixing something, and what would it be? Here is a possibility, pursued by Walley (1991). We have a language L and a body of evidence K. The probability of a sentence S, relative to evidence K, is an interval $[p, q]$. Since the lower bound for $\neg S$ is one minus the upper bound for S, probabilities are determined by the lower probability bounds. Walley deals primarily with lower bounds (actually, lower bounds of expectations), from which, of course, upper bounds are derivable as the lower bounds of negations: $\overline{P}(X) = \underline{P}(\neg X)$.

Walley proves a theorem, the convexity theorem (Walley 1991, p. 79), that shows that if $\underline{P_1}$ and $\underline{P_2}$ are lower previsions, so is their combination, defined as $\underline{P}(X) = \lambda \underline{P_1}(X) + (1 - \lambda)\underline{P_2}(X)$, and if $\underline{P_1}$ and $\underline{P_2}$ avoid sure loss, so does $\underline{P}$; and furthermore if $\underline{P_1}$ and $\underline{P_2}$ are *coherent*, so is $\underline{P}$. It follows that the set of linear previsions for which $\underline{P}$ is a lower bound is also convex. Since a linear prevision whose domain is a field is a (real valued) probability, this seems helpful.

In fact, however, it is less helpful than it seems. A lower prevision on a set of gambles is the set of maximum prices an individual is willing to pay for those gambles. It is not clear how, for a given set of gambles, there can be more than one lower prevision. Of course, two different individuals could be involved, and the import of the theorem then would be that if the individuals have previsions that avoid sure loss and that are coherent, then any linear combination of their provisions will have the same properties. This is quite different from Levi's approach under which an individual's credal state is characterized by a set of conditional probabilities; in that case, to say that the set of conditional probabilities is convex is to say something important. It is important because for Levi, as for most writers, "beliefs" come in real valued quantities, and those real valued quantities are what are used in computing expectations, and thus determining decisions. Levi's decision theory (1967) does make use of the notion of E-admissibility, which in turn depends on the convex set of conditional probabilities representing the agent's credal state. We return to this in the section after next.

Convexity for probabilities, in Walley's sense, holds also for (interval) probabilities in my sense. It is easy to see that if P_1 is a classical probability defined on a field of statements in L and K is a set of evidence statements in L such that $P_1(S) \in Prob(S, K)$, and the same is true of P_2, then the same is obviously true of $\lambda P_1(S) + (1 - \lambda)P_2$.

What may be more interesting is the question of whether there always exists a classical probability function that satisfies all our interval constraints and, even more interesting, whether there exist classical probability functions for which the lower probability of a statement S is a bounding point, and classical probability functions for which the upper probability of S is a bounding point.

The answer to the first question is yes, for any finite number of statements. The statements can be embedded in a field of statements, each of which can be expressed as a disjunction of atoms. The atoms a_i have probabilities, relative to K, of $[p_i, q_i]$. Since $\sum p_i \leq 1 \leq \sum q_i$, there is no problem about finding a probability P that satisfies the constraints on the atoms. It is a theorem of evidential probability that if $Prob(S_i, K) = [p_i, q_i]$, $Prob(S_j, K) = [p_j, q_j]$, and $Prob(S_i \vee S_j, K) = [p, q]$, then either $p_i + p_j \leq p \wedge q \leq q_i + q_j$ or $p \leq p_i + p_j \wedge q_i + q_j \leq q$. In either case there are points in $Prob(S_i, K)$ and $Prob(S_j, K)$ such that their sum is in $Prob(S_i \vee S_j, K)$.

The answer to the second question is no. It is often the case that the probability of a disjunction is more precise than the probability of either disjunct. (Consider a die that we know to be biased toward 1 and against 2, or toward 2 and against 1; the probability of getting a 1 or a 2 may still be $\frac{1}{3}$.) So suppose that $Prob(S_1, K) = [0.05, 0.60]$ and $Prob(S_2, K) = [0.10, 0.35]$, where S_1 and S_2 are logically exclusive, and $Prob(S_1 \vee S_2, K) = [.50, .60]$. Set the classical probability $P(S_1) = 0.05$. We then must have $P(S_1 \vee S_2) \in [0.5, 0.6]$. But then $P(S_2) \geq 0.45$, which violates the constraint that $P(S_2) \in [0.10, 0.35]$.

Of course none of this says anything about beliefs, even if we take beliefs to be constrained by probabilities. That the set of constraints is convex does not entail that the beliefs so constrained are convex. That is another question.

4.3. Convexity in Belief

Let us consider the set of beliefs that are not full beliefs, and whose content can be expressed in sentences of our formal language. The idea of convexity of belief as suggested by Levi (1999) and others is based on the two assumptions: First, that there are *degrees* of belief and that they are real valued, and second, that probabilities, though not precise, can be represented by *sets* of classical conditional probabilities (Levi 1999). The idea is this: A *confirmational commitment* is a *set* of conditional probability functions $Q_K(y|x)$ to which an agent with evidence K is committed (ibid., p. 513). Confirmational commitments are required to satisfy certain constraints of rationality: credal consistency, credal coherence, and credal convexity. Credal convexity is what interests us here: It stipulates that the set of permissible functions $Q_K(y|x)$ is convex – that is, that if $Q_k^1(y|x)$ is permissible and $Q_k^2(y|x)$ is permissible, then for every $\alpha \in [0, 1]$, $\alpha Q_K^1(y|x) + (1 - \alpha)Q_K^2(y|x)$ is permissible.

Note that this represents a global choice on the part of the agent; he or she cannot use one Q-function for y and a different one for $\neg y$. Both $Q_1(H|t) = 0.3$ and $Q_2(H|t) = 0.5$ may be part of the agent's confirmational commitment. (H is for heads; t is for toss.) If so, so must $Q_3(H|t) = 0.4$ be part of that commitment, by convexity. But the agent cannot adopt Q_1 for H and Q_2 for $\neg H$.

I have argued elsewhere (Kyburg 2003) that there are no degrees of belief. If there are real valued degrees of belief, and the degree of belief in y, given a hypothetical item of evidence x, of the rational agent with evidence K, is the value of some $Q_K(y|x)$ in his confirmational commitment, then although degrees of belief corresponding to other elements $Q'_K(y|x)$ of the agent's confirmational commitment might also be rational, they are not degrees of belief that the agent has. The agent's credal state consists of a collection of confirmational commitments (Levi 1999). Although Levi does not stipulate that an agent's doxastic state should be represented by a *single one* of these coherent Q functions, he does require that the expectations that underlie rational choice be computed from "a" Q function. The suggestion is that although there may be a set of Q functions that characterize an agent's credal state, these are functions that are "rationally permissible" for the agent. The agent ought to have real valued degrees of belief, and these real valued degrees ought to be represented by *one* among the Q functions.

There is another way of relating probabilities to belief. As Keynes observed (1952), probabilities are just partially ordered. In the case of our interval valued probabilities, it is natural to say that one probability is greater than another just in case every point in its interval is greater than any point in the interval corresponding to the other probability. This yields a partial order. Furthermore, it is natural to say that *beliefs* are partially ordered. However sensitive or insensitive we are to nuances of belief, we will surely acknowledge that there are some beliefs that are greater than others. My partial belief that it will rain tomorrow is pretty vague, but it is definitely *greater than* my partial belief that a large meteorite will land in my yard tomorrow night.

What we can demand of the rational agent, then, is that his or her degrees of belief conform to his or her partial probabilities in the sense that if the partial probability of x is greater than the partial probability of y, then the degree of belief in x should be greater than the degree of belief in y. Whether to demand the converse is unclear. If degrees of belief are articulated more finely than probabilities, then of course one could rationally have a degree of belief in x greater than a degree of belief in y even though there is no relation of "greater than" obtaining the probabilities of between x and y.

In the case in which we are concerned with statements whose probabilities are quite vague, this seems right. I could bet on rain tomorrow at any odds better than 2:1; and I could bet on no rain tomorrow at any odds better than 2:1; even when the probability of rain tomorrow, relative to my body of knowledge, is the large interval [0.2, 0.8]. If my betting behavior is construed as revealing my degrees of belief, it appears that my degree of belief in rain is [0, 0.3] and my belief in no rain is also [0, 0.3], yielding an interval [0.3, 0.7] that might be taken to represent my degree of belief in rain.

Where does convexity come in? If we represent amounts of belief ("degrees" seems to connote real valued magnitudes) by intervals, then convexity might be taken to mean that any point within that inteval is also a possible amount of belief. Thus if the amount of belief I have in rain is represented by the entire interval [0.3, 0.7], perhaps.

If we represent amounts of belief by intervals without gaps, then the mapping from probabilities to *rational* belief becomes quite straightforward: The amount of rational belief in a statement S, given evidence K, is exactly the same interval as the probability of S given K: $Prob(S, K)$. Convexity of belief merely stipulates that the intervals representing *rational* belief should not have gaps in them. This has no relation at all to the convexity of the sets of classical probability functions that satisfy the evidential probability constraints.

4.4. Convexity and Decision

The linear combination of any two classical probability functions that satisfy the constraints of evidential probability will also satisfy those constraints, just as, on Levi's view, the linear combination of any two Q-functions (representing confirmational commitments) in a credal state will also be in that credal state. How does this effect decision theory?

Almost anyone agrees that when you have a probability function that yields probabilities for the outcomes of your potential acts, and a utility function that yields values for the outcomes of your potential acts, then you *ought* pragmatically to choose an act that has a maximal expectation. (There are also those who seem to say that whatever act you choose does this and that your choice just reveals something about your subjective probabilities and your subjective utilities. Those we leave to one side.) When you have a set of probability functions and a set of utility functions, the recommendations of reason are not so clear.

In particular, it is easy enough to find pairs of probability distributions such that either one would lead to one course of action, but such that their combination would lead to a different course of action. A simple example consists of tossing a coin that we know to be biased so as to yield 60 percent heads, or biased so as to yield 40 percent heads. According to the evidential view of probability, the probability of heads is the interval [0.4, 0.6]. If beliefs are real valued and convex, then a perfectly reasonable "degree of belief" on heads is 0.5 (as well as 0.4, 0.6, and all the other points in between). But this would *not* be a reasonable "degree of belief" to have; in particular, if degrees of belief are tied tightly to behavior, it would not be rational to offet to sell for $.40 a ticket for which you would have to pay a dollar in case of heads. Fifty cents might be a reasonable price for such a ticket, but $500 would not be a reasonable amount for which to offer 1,000 tickets that would pay a dollar if heads on 1,000 tosses.

One way of accounting for this is to note that the tosses are not represented as a binomial variable with a parameter of 0.5 in any case: The parameter is either 0.4 or 0.6. To consider the rationality of believing that between 450 and 550 of the next 1,000 tosses yield heads, we must take account of conditioning.

5. CONDITIONING

The problem with convexity becomes even worse once we consider conditioning. In the example we just brought up, in which we know of a coin that it is either biased so as to yield heads 40 percent of the time or 60 percent of the time, we can be almost certain that the coin will land heads in a large set of tosses nearly 40 percent of the time or that the coin will land heads in a large set of tosses nearly 60 percent of the time. But whatever kind of coin we have, of course, we can also be practically certain that the coin will not land heads 50 percent of the time.

Why is this a difficulty? Because we expect relative frequencies to reflect probabilities; if 0.50 is an acceptable probability for heads, then 50 percent should be an acceptable probability for the relative frequency of heads, other things being equal. But in this case it is not; we can be almost certain that the relative frequency of heads is not close to 0.50.

Things get even worse when we consider coins about which we know very little. Suppose all we know of a coin is that it yields between 1 percent and 99 percent heads in a binomial distribution. Now 50 percent is possible, and since a parameter of 0.5 corresponds to the maximal variance, we can say that the odds are greater that in the long run 50 percent of the tosses will yield heads than that we will get any other frequency.

Note that we are not stipulating a "flat" prior; we are trying not to stipulate any prior by using a collection of priors. But whatever priors we use, we must also be prepared to use linear mixtures of them. As in the case of the two kinds of coins, these linear mixtures may include cases we don't want.

It may well be claimed that we are confounding two kinds of things: what we have reason to believe about the long run behavior of the coin and what we "believe" about the next toss. That is just my point. What we have reason to believe about the long run behavior of the coin is what should constrain our beliefs concerning the next toss. If what we have reason to believe is vague, so should be those constraints. If those constraints are vague, then it is not clear how to apply conditioning.

My view, as already suggested in section 2.2, is that conditioning makes sense only when applied to frequencies. Beliefs are constrained not by conditioning, which makes sense only when we have real numbers to work with, but by probabilities relative to a corpus of evidential certainties. This corpus may yield (vague) probabilities that are based on conditioning – for example, when the coin we have been talking about comes from a bag of coins with a known (or approximately known) ratio of 40 percent coins to 60 percent coins – but that is quite different from conditioning on real valued degrees of belief.

6. CONCLUSION

In conclusion, if we do not have real valued degrees of belief, it is not clear what convexity comes to. At best it is a reflection of the convexity of the set of probability functions that conform to the probability intervals that constrain vague beliefs. But in that case convexity undermines the possibility of conditioning. How we get from evidence to new beliefs is by adding the evidence to K (in some sense of "add") and taking the new interval valued probability function to produce new constraints on belief. It is only rarely that this process can be represented by conditioning; it is rare that an interval of belief can be obtained from other intervals of belief by a process of division, if division makes sense at all.

REFERENCES

Keynes, John Maynard. 1952. *A Treatise on Probability*. London: Macmillan.

Kyburg, Henry E., Jr. 1990. "Theories as Mere Conventions." In Wade Savage (ed.), *Scientific Theories*, pp. 158–74. Minneapolis: University of Minnesota Press.

Kyburg, Henry E., Jr. 2003. "Are There Degrees of Belief?" *Journal of Applied Logic* 1: 139–49.

Kyburg, Henry E., Jr., and Choh Man Teng. 2001. *Uncertain Inference*. New York: Cambridge University Press.

Levi, Isaac. 1967. *Gambling with Truth*. New York: Knopf.

Levi, Isaac. 1999. "Value Commitments, Value Conflicts, and the Separability of Belief and Value." *Philosophy of Science* 60:509–33.

Ramsey, F. P. 1931. *The Foundations of Mathematics and Other Essay*. New York: Humanities Press.

van Fraassen, Bas. 1984. "Belief and the Will." *Journal of Philosophy* 81: 235–56.

Walley, Peter. 1991. *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.

8

Levi's Chances

D. H. Mellor

Isaac Levi and I are old friends who have argued for decades about philosophical topics that interest us. Although – or perhaps because – we often disagree, I have learned more from our debates than I sometimes admit. So I was especially pleased to be asked to contribute to this volume, and what follows is offered with respectful affection if not with much hope of inducing complete agreement.

Levi (1977, pp. 186–7) stresses "the fundamental importance . . . to the understanding of the conception of chance . . . of providing an account of direct inference," as opposed to "the gratuitous, diversionary and obscurantist character of such 'interpretations'" as von Mises's (1957) frequency theory and my (1971) and other propensity theories. Levi's own theory of chance (1980, chs. 11–12) amply meets his own desideratum. In doing so, however, it differs less than he thinks from its rivals.

1. DIRECT INFERENCE

By "direct inference" Levi means a principle "which stipulates how knowledge of chances . . . determines credal judgments about the outcomes of trials on chance setups" (1980, p. 86), where "credal judgments" means what he calls credal probabilities, which for brevity I call *credences*, "to be used in practical deliberation and scientific inquiry in computing expectations" (1977, p. 165). He illustrates his principle as follows.

Suppose *X* knows the following bits of information:

(i) The chance of coin *a* landing heads on a toss [of kind *S*] is .5 and of landing tails is also .5.
(ii) Coin *a* is tossed at *t*.
(iii) The toss of *a* at *t* is also of kind *T*.

. . . [Then] knowledge of (i), (ii), (iii) and that the information that the toss is of kind *T* is stochastically irrelevant . . . warrants assigning the hypotheses that the coin *a* lands heads at *t* and that the coin *a* lands tails at *t* equal [credences] of .5. (Levi 1980, pp. 251–2)

The basic idea of this principle has long been widely accepted. As Hacking (1965) remarked of the version that he called the "frequency principle" and said was trivial, it

> seems so universally to be accepted that it is hardly ever stated . . . [that] if all we know is that the chance of *E* on trials of kind *K* is *p*, then our knowledge supports to degree *p* the proposition that *E* will occur on some designated trial of kind *K*. (p. 135)

Stated like this, however, the principle is all but useless, since rarely if ever is the fact that – in Levi's example – a coin *a*'s chance of landing heads on a toss of kind *S* is 0.5 *all* we know that might or should affect our credence in *a*'s landing heads. Hence Levi's requirement that we also know it to be *stochastically irrelevant* that the toss is of some other kind *T*; hence also the italicized caution in my (1971) statement that

> some personal probabilities can be made more reasonable than others in a person *suitably situated* by his being aware of a corresponding objective probability (p. 29; emphasis added)

and the admissibility condition on *X* in Lewis's (1980) "principal principle," that if we

> let *C* be any reasonable initial credence function[,] . . . *t* be any time[,] . . . *x* be any real number in the unit interval[,] . . . *X* be the proposition that the chance, at time *t*, of *A*'s holding equals *x* [and] *E* be any proposition compatible with *X* that is *admissible* at time *t*, then $C(A/XE) = x$. (p. 87; emphasis added)

As these quotations show, defenders of direct inference recognize the need to qualify its basic idea, and to do so without making their principles inapplicable or trivial. The devil, as always, is in the detail, and Levi's details are as devil-free as most. Still, differences of detail, while important, should not obscure the common basis of all these principles, and the shared conviction of their advocates that, as Levi says, "an account of [them is of fundamental importance] to the understanding of the conception of chance" – whether or not, as Lewis claims (1980, p. 86), they "capture all we know about chance."

But suppose a principle of direct inference did tell us all we need to know about chance. It might still not tell us all we need to know about how evidence should affect our credences. Levi, however, thinks it does, since he thinks that an "objectivist inductive logic . . . restricted to credal coherence and direct inference . . . is a complete inductive logic," despite the fact that it is, as he says, "insufficient for the purposes of the Jeffreys-Carnap program" (1980, p. 87; Jeffreys 1961; Carnap 1962). Still, one may reject that ambitious program for inductive logic while wishing to replace or supplement direct inference with some Bayesian or other principles (see, e.g., Jeffrey 1983; Howson and Urbach 1993). And even if, as I and Levi believe, direct inference cannot be replaced or reduced to anything else, it may still need supplementing by principles that apply when it does not, that is, when we know no relevant chances. In other words, even if Levi's inductive logic is *correct*, it may not be *complete*. But that is a matter for another time, since all that concerns me here is chance, and what the fact that some principle of direct inference is justified can tell us about it.

2. CONDITIONALS AND CREDENCES

Two other matters on which Levi and I differ need also not detain us long, despite their role in his account of direct inference. One is about conditionals such as "if coin *a* were tossed 1,000 times it would land heads approximately 500 times," which Levi says is supported by "*a* is a fair coin" (1980, p. 276), just as something such as "object *o* would dissolve if put in water" is supported by "*o* is soluble." The difference here is that Levi denies, and I assert, that these conditionals have truth values, which is why he says they are *supported* by statements that I say *entail* them (Levi 2002; Mellor 2003, sec. 9). However, as what really matters here is *what* conditionals statements such as "*a* is a fair coin" or "*o* is soluble" support or entail, not whether they have truth values, I here use Levi's term "support."

In Levi's coin case, I agree that "*a* is a fair coin" does support (i) "if *a* were tossed 1,000 times it would land heads approximately 500 times," albeit for reasons that I know he will not accept. Levi notes that

> students of possible-world semantics will [deny this and] suggest instead that the chance statement supports the judgment that [ii] if *a* were tossed 1,000 times *in all probability* it would land heads approximately 500 times.
>
> (p. 276, emphasis added)

This is a replacement that he rejects. But I think "*a* is a fair coin" supports (i) *because* it supports (ii): Since I take "in all probability" to imply that *a*'s chance of landing heads approximately 500 times in 1,000 tosses is close enough to one for direct inference to warrant a credence in that outcome that is so close to one that I say (although Levi would not) that it amounts to full belief.

Our other irrelevant disagreement is about whether ascriptions of credences are normative or factual. Like most philosophers, we agree that, as Levi puts it,

> [t]he rationale for credal coherence is found in the account of how [credences] function in deliberation and inquiry in the evaluation of feasible options with respect to expected utility. (1980, p. 261)

Where I part company with him and most others is that I, like Ramsey (1926), take this rationale to be part of a *descriptive* rather than a normative decision theory (Mellor 2005). Thus I think the fact that, for any proposition A, the values of my credences in A and in not-A must add up to one, is a theoretical idealization rather than a requirement of rationality. But this difference, although serious elsewhere, is irrelevant here, where all that matters is that principles of direct inference, which relate credences to chances, are normative, which I agree they are. What they tell us, rightly or wrongly, is when we *should*, not when we *do*, equate our credences to chances we know. The question is why we should do what these principles tell us to do: When direct inference is right, what makes it so?

3. CREDENCES AND FREQUENCIES

Let us start by asking what makes some credences more useful than others. As my last quotation from Levi tacitly implies, he agrees that our credences should satisfy a theory that tells us to maximize expected utility. Why should they do this? The answer given by Ramsey (1926), who first used such a theory to define credences (which he calls "degrees of belief," a reading that Levi rejects for reasons that make no odds to what follows), is that

> the very idea of partial belief [in a proposition A] involves reference to a hypothetical or ideal frequency; supposing goods to be additive, belief of degree *m*/*n* [in A] is the sort of belief which leads to the action which would be best if repeated *n* times in *m* of which the proposition [A] is true. (p. 84)

Thus suppose, for example, my credence $\frac{1}{2}$ in a coin *a* landing heads on tosses of kind *S* makes me bet repeatedly on this result but only at evens or better. This is the action that is best for me if $\frac{1}{2}$ of the tosses I bet on land heads; for then to bet on heads at any shorter odds would lose me money. More seriously, suppose I am an insurer with a credence *p* that any one person of a different kind *S* – defined by their age, sex, health, occupation, place of residence, and so on – will die within a year. Then for each unit (pounds, euros, etc.) of life insurance that I offer to people of kind *S*, I will set a basic annual premium (i.e., before covering overheads and profit) of at least *p* units, for this is the least that will stop me losing money if *pn* of every *n* such people who buy my insurance die within a year.

Ideally, then, our credences should equal the relevant frequencies: of coin tosses of some kind *S* that land heads; of people of some other kind *S* who die within a year; and in general of instances of some kind *S* that are also of a specific kind *R*. But this is no use as a prescription, for if we knew in advance how coin tosses will land or when people will die, our credences in those events could all be 0 or 1 and we could act on *actual* rather than on merely *expected* utilities. In real life, however, we need credences other than 0 or 1 based on something that (1) we *can* know in advance but (2) will also give us some assurance that they will equal, or at least be close to, the actual frequencies that would make them the right credences to have. So any inductive logic that, like Levi's, sometimes tells us to equate credences with chances needs a theory of what chances are that (1) makes them knowable in advance and (2) relates them to frequencies in a way that explains why our credences should equal them. What are the options?

4. CHANCES AS FREQUENCIES

The main theories of what chance is are too well known to need stating in much detail here. Nor need we look at all their variants, most of whose distinguishing features are irrelevant for present purposes. For these purposes we may also set aside Bayesian and other subjective theories that seek to reduce chances to (e.g.) "resilient" credences (Skyrms 1980, part I), since they would make direct inference redundant. The only theories of chance we need look at here are *frequency* and *propensity* theories.

First, frequency theories, starting with the finite frequency theory (Russell 1948, part V, ch. III), which identifies the chance of an *S* being *R* with the relative frequency of *R*s in the finite reference class of all actual *S*s, such as the fraction of all people of kind *S* who die within a year. There is of course more to it than this, since no one who thinks that all chances are frequencies thinks that all frequencies are chances. No one, for example, will take the fraction of Scots born in July 1976 or May 1960 who die in China to be a chance. Anyone who thinks, for some *S* and *R*, that the fraction of *S*s that are *R* is a chance will think that there is something like a law-like link between being *S* and being *R*. What this means, and when and why we should believe it, are indeed good questions that however for present purposes we need only assume have some tenable answers.

Provided, then, that the finite frequency theory can somehow distinguish frequencies that are not chances from frequencies that are, it can easily make its chances meet condition (2) above. If our credence that an *S* will be *R* should equal the fraction of *S*s that are *R*, and the chance of *S*s being *R* just *is* that fraction, then direct inference is undoubtedly justified. Unfortunately, as our examples show, it is also useless, since the fraction we need will not be known in advance. Theories that take chances to be actual finite frequencies will therefore not support an inductive logic that is confined, as Levi's is, to coherence and direct inference. They will also need principles of statistical inference from observed frequencies of *R*s in past samples of *S*s (e.g., last year's death rates among people of kind *S*) to the frequency of *R*s in the total population of *S*s (all people of kind *S*) that they identify with an *S*'s chance of being *R*. But this makes a direct inference from that chance to a credence that an *S* is *R* both trivial and dispensable: For as our inference from an observed sample to the total population is now doing all the epistemic work, we may as well treat it as an inference direct from the sample to a credence (or interval of credences), thereby cutting out chance and direct inference altogether.

And as for a finite frequency theory, so at first sight for the limiting frequency theories of Reichenbach (1949), von Mises (1957), and others, as applied to infinite populations. For if we cannot know frequencies in finite reference classes in advance, we can hardly know the limits, if any, of sequences of frequencies in finite subclasses of infinite reference classes. However, limiting frequency theories are usually applied not to *actual* infinite reference classes but to *hypothetical* ones. One reason for this is a well-known objection to the finite frequency theory, namely, that it makes the *possible* values of an *S*'s chance of being *R* depend on how many *S*s there happen to be. Thus if there are only four actual *S*s, the theory limits their possible chances of being *R* to 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, and 1, which is absurd. Nor do larger finite numbers of actual *S*s improve the situation much. For however many *S*s there are, the finite frequency theory will still rule out infinitely many possible values of an *S*'s chance of being *R*. *Hypothetical* limiting frequency theories overcome this limitation by identifying chances with what the corresponding limiting frequency *would* be if the relevant reference class were infinite. This makes a coin *a*'s chance of landing heads on tosses of kind *S* the limiting frequency of heads in an infinite class of hypothetical tosses of this kind, and similarly in other cases.

Yet this development seems to make it harder still, if that were possible, to know chances in advance. If we cannot know actual finite or limiting relative frequencies in advance, how can we know merely hypothetical ones, which not being actual can never actually be observed? Oddly enough, it can be done; but to do it we must abandon frequency theories of chance for a propensity theory.

Before showing how a propensity theory explains our ability to know chances in advance, I need to make a point about conditionals on which I and Levi partly agree, in substance if not in terminology. The point is about conditionals that, like "object *o* would dissolve if put in water," are related to disposition statements such as "*o* is soluble." Levi and I differ here in that I see more merit than he does in a possible-world semantics for these conditionals. But even if that semantics tells us what such conditionals *mean*, I agree with Levi that it does not tell us what makes them *true* (as I put it) or *supports* them (as he puts it). Setting aside questions of *how* soluble something must be to be soluble, and what to say about objects whose solubility is affected by their being put in water, we agreed in section 2 that what supports something such as "*o* would dissolve if put in water" is "*o* is soluble" or rather – since the mere proposition that *o* is soluble does not on its own support anything – the fact that "*o* is soluble" is *true*. It is truths like this about the actual world, not truths about other possible worlds, that determine which contingent conditionals are supported in our world.

Next, we can use this link between conditionals and disposition statements to say, without explicitly invoking other possible worlds, what predicates such as "soluble" mean. For whatever we think of possible world semantics, no one will deny that we can understand "*o* would dissolve if put in water" without knowing the meaning of "soluble." But then we can use that conditional to say what it means to call any object *o* soluble: roughly, that *o* would dissolve if put in water. More precisely – taking "soluble" to mean soluble in water and still not saying how soluble an object must be to be soluble – for any object to be soluble at any time *t* is for it to dissolve if it is put in water at *t and remains soluble*. The italicized proviso is needed to cover the possibility mentioned above, that putting a soluble object in water may make it *in*soluble.

Yet how can a conditional with this proviso tell us what "soluble" means? For if "soluble" occurs *within* the conditional, as it does, must we not know what it means in order to understand the conditional? That is true, but only up to a point, the point being that it stops us replacing the predicate "is soluble" with a predicate "is such that _" which does not contain "soluble." Nevertheless, the conditional can still be used to introduce or explain this predicate, since all that anyone who understands conditionals of this kind needs to know to understand "soluble" is that, by definition, any soluble object would dissolve if it was put in water and remained soluble.

Similarly, on a propensity theory, for chance: "*o* is soluble" and "*a* has chance *p* of landing heads" have similar explanatory links to conditionals.

The conditional in the case of chance is like the one that hypothetical limiting frequency theorists use to say what *a*'s chance of landing heads *is*: namely, what the limiting frequency of heads would be if *a* were tossed endlessly. The crucial difference is that I and other propensity theorists do not *identify* actual chances with merely hypothetical frequencies. That seems to us as mistaken as identifying *o*'s actual solubility with its merely hypothetical dissolution in water. The mistake, however, is easily made, since it arises from the use of quantitative conditionals to provide measures of the quantities they define. Thus, just as we can measure *o*'s actual solubility by how much of it would dissolve in a liter of water, so we can measure *a*'s actual chance of landing heads by what the limiting frequency of heads would be if *a* were tossed endlessly. This convenient convention generates a trivial equality of values – between *o*'s solubility and how much of it dissolves in a liter of water, and between chances and limiting frequencies – which is then easily mistaken for an identity of the quantities that have these values.

On a propensity theory, then, chance is linked to hypothetical limiting frequency not by identity but by a link like that between *o*'s solubility and its hypothetical dissolution. That is, *a*'s having a chance *p* of landing heads supports something like "if *a* were endlessly tossed, the limiting frequency of heads would be *p*." More precisely, it supports "if *a* were endlessly tossed *with each toss having a chance p of landing heads*, the limiting frequency of heads would be *p*." Here the italicized proviso covers the possibility that tossing *a* repeatedly might change its chance of landing heads, that is, that repeated tosses might not be physically *independent*, which is the analogue of a soluble object being made insoluble by being put in water.

But then, by analogy with solubility, may we not use this conditional to tell us what chance predicates mean? We can of course no more use a conditional with the above proviso to *replace* a chance predicate with one that contains no such predicate than we could in the case of solubility; but neither I nor Levi ever thought we could. Even so, as with solubility, the conditional might still serve to introduce or explain the chance predicate – provided that all that anyone who already understands conditionals of this kind needs to learn in order to understand "has a chance *p* of landing heads" is that, by definition, an endless sequence of tosses with that chance would have a limiting frequency *p* of heads. But is that so?

The answer is "not really." The link between chance and frequency is more complex than that between solubility and dissolution. To understand chances we need to know how they support the conditionals about hypothetical *finite* frequencies that entail the existence of limiting ones. Specifically, we need to know what makes chances satisfy the *law of large numbers*. This says that, if a has a chance of landing heads on any one toss that is independent of the result of any other toss, then for any positive δ and ε, however small, there is a p and an n such that, if a were tossed n or more times, the chance of f_n, the frequency of heads, lying within δ of p would be within ε of 1. This makes the number p the hypothetical limiting frequency that provides our measure of a's actual chance of landing heads on any one toss. The existence of a limit so defined is what, on a propensity theory, it *is* for a to have a chance of landing heads; and that is how the conditionals used to state the law of large numbers can tell us what chances are.

(But what then of the objectivity of chances, if all that links them to hypothetical frequencies is the law of large numbers? It is after all well known that our *credences* in facts about how often a would land heads on many hypothetical tosses also satisfy this law, provided we take the tosses to be *exchangeable* in the sense of de Finetti (1937). However, this only means that taking these tosses to be exchangeable will make my credence p in a's landing heads on any one toss give me a high credence that the frequency of heads on many tosses will be close to p. That is irrelevant to what concerns us here, which is how to derive credences from chances by direct inference. All that is relevant here is *why* we take these hypothetical tosses to be exchangeable: Namely, because, by hypothesis, we take them to be *objectively* independent, that is, take each toss to have a chance p of landing heads that is physically unaffected by the result of any other toss. It is this, plus direct inference, that makes us take these tosses to be exchangeable, thus making our credences in facts about the frequencies of heads also satisfy the law of large numbers.)

6. PROPENSITIES AND CREDENCES

How well does a propensity theory of chance, as sketched above, explain the applicability and validity of direct inference? That is, how well does it meet the two desiderata of section 3: that a theory of what chances are should (1) make them knowable in advance and (2) relate them to frequencies in a way that explains why we should make our credences equal them? I have said that a propensity theory meets condition (1) better than frequency theories do, a claim I have not yet made good. But before trying to do so, I must show that a propensity theory can also meet condition (2).

Given the ideal link between credences and frequencies stated in section 3, condition (2) would be met perfectly only if a chance p of heads on each of n independent tosses entailed that precisely pn of them would land heads. But that, as we have seen, is impossible, and to ask it is to cry for the moon. What we *can* know, given the law of large numbers, is that p is the only value to which the frequency f_n of heads in n tosses has a high and increasing chance of being close as n increases. That makes p the credence in heads to which, for all n, any actual frequency f_n of heads on n tosses has the best chance of being as close as would make no odds to any decisions we used it to make, and therefore of being the most useful value of this credence. This I believe is enough to justify Levi's principle of direct inference from known chances to credences.

There is, however, a well-known objection to this argument. The objection is to its using facts about the chances of outcomes of many tosses to justify a credence in the outcome of a single toss. For why should the fact that a credence would (probably) serve me well if it fixed the odds I would accept for *repeated* bets on (e.g.) coin tosses landing heads make it the right credence to have for a *single* bet that I need have no intention of repeating? The answer to this rhetorical question lies in the fact that, as the decision theory we use to define credences recognizes, no credence is limited to informing only one decision. Once acquired, each of our credences combines with many different possible utilities to fix the expected utilities of many possible actions that we never do, either because we lack the utilities that would make them worth considering or because, even with our actual utilities, their expected utilities are less than those of some alternatives. This being so, a credence's justification depends not just on the few decisions to which it actually leads but on all the other decisions to which, with different desires, it would have led. That is what enables the chances of facts about the frequencies of heads on many merely possible coin tosses to justify a credence in a single actual toss landing heads.

7. CHANCES AS PLACEHOLDERS

So much for condition (2) on a credible theory of what chance is: The law of large numbers enables a propensity theory to meet it as well as it can be met, and a fortiori as well as any frequency theory can. What then of condition (1), that chances be knowable in advance, which I said at the end of section 4 defeats the hypothetical limiting frequency theory? Yet how, if it does so, can a propensity theory survive, when it says that chances get their values from the hypothetical limiting frequencies whose existence they entail? For how then can we know the former in advance when we cannot know the latter?

To see how, we must first see how the propensity theory answers another question that we noted in section 4 faces all frequency theories: that of saying for *which* R and S the frequencies (or their limits) of Rs in a reference class of

*S*s is a chance. Now when a reference class has only actual members, we can always specify it as the class of all actual *S*s. Setting aside any vagueness in what the predicates "*R*" and "*S*" apply to, this ensures that, whether or not the frequency of *R*s in this class (or its limit, if any) *is* a chance, it will at least have a definite value. Not so if the reference class contains infinitely many merely possible members. Then the very existence of a limiting frequency of *R*s in a reference class of *S*s requires *each S* to have a property, which propensity theorists say *is* its chance of being *R*, that makes all finite classes of actual or hypothetical *S*s satisfy the law of large numbers. It is this property that ensures that the infinite reference class of hypothetical *S*s *has* a limiting frequency of *R*s with whose value an *S*'s chance of being *R* can then be equated.

This is why propensity theorists never face the question of which frequencies are, or rather correspond to, chances; for on their theory, the limiting frequencies that correspond to chances will not exist unless the chances do. The questions this theory faces instead are how to tell, for any given *R* and *S*, (a) that *S*s *have* a chance of being *R* and (b) what value this chance has. And those, as we shall now see, are the questions whose answers enable the theory to say how, as our condition (1) requires, we can know chances in advance.

While I have said that the propensity theory credits each *S* with a property that the theory identifies with an *S*'s chance of being *R*, I have not said that I take properties to be universals, sets of resembling tropes or particulars, or something else again. But although this is a question that I take seriously and Levi does not, it is fortunately not one that I need to answer here. All I need here are two assumptions that both Levi and I do accept: first, that statements such as "*a* has a chance *p* of landing heads" resemble disposition statements such as "*o* is soluble" in having truth values; and second, that the predicates that occur in both these statements are what Levi and Morgenbesser (1964) call "placeholders." The following quotation illustrates what Levi means by this:

> Some iron bars attract iron filings placed near them and others do not. As a first step toward understanding the differences between the two sorts of iron bars, *X* may say that one sort of bar has a disposition to attract iron filings and the other does not. Of course, this description of the difference is but a first step. That is why explicit disposition predicates are placeholders for more adequate characterizations of the relevant differences. Nevertheless, they have an important function, and in many instances, an indispensable one, in inquiry and deliberation. (Levi 1980, p. 268)

Similarly with chance predicates. For now suppose, with Levi, that *X* has credences 0.5 and 0.9, respectively, in coins *a* and *b* landing heads if tossed, and the credences that follow from these in *a* and *b* landing heads *r* times on *n* tosses that *X* takes to be exchangeable. Then, as Levi says,

> there would be wide agreement that such a credal state makes no sense unless there is some significant difference in the characteristics of coin *a* and coin *b*.
>
> That is not to say that *X* should be in a position to offer an explanatorily adequate characterization of the difference between the coins; but he should be committed to the view that there is a difference in traits. The coin *a* has some property *C* such that given knowledge that an object has *C*, ceteris paribus, *X*'s credal state for hypotheses specifying relative frequencies of heads on *n* tosses should be as specified above. Similarly, *b* has some property *C′* knowledge of the presence of which licenses a credal state of the sort attributed to hypotheses about *b*'s behavior.
>
> One way of putting it is to say that coin *a* is unbiased whereas coin *b* is heavily biased in favor of heads. Another way to put it is to specify the explicit chance predicates that are true of *a* and of *b* concerning outcomes of *n* tosses. (Ibid., 269)

Levi then admits that describing this difference between *a* and *b* as a difference in their chances of landing heads is "deficient," just as it is to describe the difference between objects that dissolve in water and objects that do not as a difference in solubility. For neither description is anything more than a placeholder for a "more adequate characterization of the relevant differences," which in the case of chance would be provided

> by integrating chance predicates into theories through inquiry as is attempted in genetics, statistical mechanics and quantum mechanics in different ways. (Ibid., 269)

The way the scientific theories that Levi invokes do this is by postulating what, in the case of dispositions, Armstrong (1993, ch. 6.VI) and others call their *categorical bases*, such as the molecular structures that distinguish soluble objects from insoluble ones. Similarly with chances, whose categorical bases are *non*chance properties, such as those that distinguish biased coins like *b* from unbiased ones like *a*. These are the properties that really explain why coin *b* tends to lands heads more often than coin *a* does.

This distinction, between these chances and their categorical bases, is what explains how we can know *a*'s and *b*'s chances of landing heads in advance, thus enabling direct inference to tell us what credences to have in their doing so. It does this by dividing our epistemic task into two parts. The hard part is testing a statistical theory of coin tossing that will say, for example, that coins with one specified nonchance property *C* have a 0.5 chance of landing heads, while coins with another such property *C′* have a 0.9 chance of doing so. We do this by using standard techniques of statistical inference to test the theory against data showing how often coins with properties *C* and *C′* respectively land heads when tossed. To establish such a theory in this way is to show that C and C' are indeed categorical bases of these two chances of coins landing heads. (These chances may of course have different bases in coins of different kinds, just as solubility and insolubility may have different molecular bases in objects of different kinds.) That is the first part of our task, which need not involve either coin a or coin b.

Once some such theory has been established, the second part of our epistemic task is easy. Before tossing either a or b we discover quite independently that a has the nonchance property C and b has the nonchance property C'. This, together with our theory, tells us *in advance* that a's and b's respective chances of landing heads are 0.5 and 0.9, thus enabling direct inference to license, in advance, those very credences. That I am sure is how Levi thinks we apply direct inference in this case, if only because I can see no other way of applying it that is consistent with what he says. I am, however, less sure that he realizes that this way of applying it presupposes a metaphysical theory – a propensity theory – of what chances are; for if he did, he could hardly take the development and defense of such a theory to be the gratuitous, diversionary, and obscurantist activity that he says it is.

REFERENCES

Armstrong, David M. 1993. *A Materialist Theory of the Mind*, revised ed. New York: Routledge.

Carnap, Rudolf. 1962. *Logical Foundations of Probability*, 2nd ed. Chicago: University of Chicago Press.

de Finetti, Bruno. 1937. "Foresight: Its Logical Laws, Its Subjective Sources." In Henry E. Kyburg, Jr., and Howard E. Smokler (eds.), *Studies in Subjective Probability*, pp. 93–158. New York: Wiley.

Hacking, Ian. 1965. *Logic of Statistical Inference*. Cambridge: Cambridge University Press.

Howson, Colin, and Peter Urbach. 1993. *Scientific Reasoning: The Bayesian Approach*, 2nd ed. La Salle, Ill.: Open Court.

Jeffrey, Richard C. 1983. *The Logic of Decision*, 2nd ed. Chicago: University of Chicago Press.

Jeffreys, Harold. 1961. *Theory of Probability*, 3rd ed. Oxford: Oxford University Press.

Levi, Isaac. 1977. "Subjunctives, Dispositions and Chances." In Levi, *Decisions and Revisions*, pp. 165–91. Cambridge: Cambridge University Press.

Levi, Isaac. 1980. *The Enterprise of Knowledge*. Cambridge, Mass.: MIT Press.

Levi, Isaac. 2002. "Dispositions and Conditionals." In Hallvard Lillehammer and Gonzalo Rodriguez-Pereyra (eds.), *Real Metaphysics*, pp. 137–53. London: Routledge.

Levi, Isaac, and Sydney Morgenbesser. 1964. "Belief and Disposition." *American Philosophical Quarterly* 1: 221–32.

Lewis, David K. 1980. "A Subjectivist's Guide to Objective Chance." In Lewis, *Philosophical Papers*, volume 2, pp. 83–113. Oxford: Oxford University Press.

Mellor, D. H. 1971. *The Matter of Chance*. Cambridge: Cambridge University Press.

Mellor, D. H. 2003. "*Real Metaphysics*: Replies." In Hallvard Lillehammer and Gonzalo Rodriguez-Pereyra (eds.), *Real Metaphysics*, pp. 212–38. London: Routledge.

Mellor, D. H. 2005. "What Does Subjective Decision Theory Tell Us." In Hallvard Lillehammer and D. H. Mellor (eds.), *Ramsey's Legacy*, pp. 137–148. Oxford: Oxford University Press.

von Mises, Richard. 1957. *Probability, Statistics and Truth*, 2nd English ed. London: Allen & Unwin.

Ramsey, F. P. 1926. "Truth and Probability." In Ramsey, *Philosophical Papers*, ed. D. H. Mellor, pp. 52–109. Cambridge: Cambridge University Press.

Reichenbach, Hans. 1949. *The Theory of Probability*. Berkeley: University of California Press.

Russell, Bertrand. 1948. *Human Knowledge*: *Its Scope and Limits*. London: Allen & Unwin.

Skyrms, Brian. 1980. *Causal Necessity*. New Haven: Yale University Press.

9

Isaac Levi's Potentially Surprising Epistemological Picture

Wolfgang Spohn

1. A BRIEF LOOK FORTY YEARS BACK

I certainly have no doubt about how much Isaac Levi has taught us in epistemological matters. Yet not for the first time I wonder at how philosophers can be so close and so different at the same time. The better the comparability, I assume, the sharper the comparison. Hence, this paper is devoted to a brief comparison of Levi's epistemological picture and mine. It mainly stays on an informal level, aiming at the basic features. The intention is not a critical one. It is rather to promote mutual understanding, because the similarities and differences are not easy to grasp.

Current discussions in formal epistemology tend to be quite specialized, building on rich but not well reflected presuppositions. So, a brief look at its history may be healthy. It is indeed necessary for understanding Levi's role and position.

The history of formal epistemology[1] is predominantly that of probability theory. Mathematical probability was the only clear structure that emerged through the centuries. Not that there would have been no alternatives at all. There are various ideas, perhaps subsumable under the heading "Baconian probability," which hung around for centuries as well. But they never took a clear and determinate shape, and thus probability theory could develop its unrivaled power, culminating, as far as its philosophical use is concerned, in the work of Frank Ramsey, Bruno de Finetti, Leonard Savage, Rudolf Carnap, and others.

[1] There can be no doubt about the lasting importance of formal epistemology. It may be the lesser part of epistemology. Still, the major part is using its central notions, such as "inference," "justification," and "reason," *without any adequate theory* in reserve. Not that formal epistemology would make unanimous offers here. But it is making offers at least – and is therefore desperately needed.

Twentieth-century philosophy of science made renewed attempts at alternatives.[2] Popper developed his account of corroboration, Hempel heroically engaged in qualitative confirmation theory. However, such attempts were not well received. So, the situation at the beginning of the 1960s was characterized by the dominance of (subjective) probability theory, the impotence of alternatives, and yet the remaining feeling that probability cannot be everything. This feeling had at least three sources: The search for alternatives, even if it had failed so far, was at the same time an expression of a sense of incompleteness. The observation that probability theory was the hobby only of formal epistemologists, but hardly used by all the other ones, was at least irritating. And the rise of doxastic and epistemic logic (see in particular Hintikka 1962) had shown that at least an elementary account of belief and knowledge besides probability was rigorously possible. All three sources pointed into the same direction: What was missing was a full-blown formal account of belief or acceptance (and, possibly, knowledge).

In my retrospective, the situation was essentially dramatized by the so-called lottery paradox, introduced by Kyburg (1961, p. 197). It expressed an already sharpened view of the situation, and its centrality soon became obvious. This was a crucial time at which formal epistemology was definitely raised to a new stage and quality. Since then, the developments have been breath-taking, but I locate their starting point rather there than anywhere else.

For sure, Isaac Levi was one of the central players in this epistemological revolution. His first book (1967) was his response to this situation. In retrospect, it has become clear how strongly his book has determined his research agenda till the present day. This is why I say one can understand his work only by looking at that situation in the 1960s.

The beauty of the lottery paradox lies in its perfect simplicity and in its surprising capacity to crystallize various basic options in epistemology. The immediate result of the lottery paradox was that there is no simple analysis of belief or acceptance in terms of subjective probability. The immediate question it raised was: What, then, is the relation between belief and probability, two obviously fundamental epistemological notions? The range of options is surprisingly similar to that concerning the relation of the mental and the physical.

One position is *eliminativism*: If belief is not analyzable in probabilistic terms, just drop it! We need not talk of anything besides probability in

[2] Did I forget formal logic starting in late nineteenth century? Of course, its import was overwhelming and still is. But deduction is not even half of epistemology. So, logic is auxiliary to epistemology, as it is to many other disciplines, rather than formal epistemology in itself.

epistemology. This is the position Richard Jeffrey, one of the other central players, ably defended throughout his life (see, e.g., Jeffrey 1992). He called it radical probabilism. In a way, it is not as radical as eliminativism in the philosophy of mind, which, in effect, is an unredeemed check on the future. Jeffrey's eliminativism is conservative, since it is a philosophically conscientious defense of the previous status quo in formal epistemology. Without doubt, as far as theoretical unity and elegance are concerned, Jeffrey's position is superior to all others. I was always attracted by this elegance.

Still, eliminativism is deeply incredible, also in epistemology. It cannot be simply mistaken or confused or superfluous to talk of belief. What, then, are the other options?

As in the philosophy of mind, the prima facie most attractive option is *reductionism*, of course. But it fares differently. Simple realizations of that option are barred by the lottery paradox.[3] And we have not seen any more involved realization of any plausibility. Reductionism in the epistemological case does not seem to be a live option.

Reverse eliminativism or reductionism is even more unfeasible; to try to get rid of probability in favor of belief is obviously crazy. Hence, any kind of monism concerning belief and probability seems excluded. Maybe both can be reduced to, or replaced by, a third thing. There are even ideas what that third thing might be.[4] However, this is not the place to expand discussion into that direction. What remains is epistemological *dualism* (or pluralism, if the need for further basic doxastic structures should be imperative).

Dualism may take different forms. One is *interactionism*, as one may call it. This consists in an integral picture, which give both belief and probability their due place and describes how they interact without reducing one to the other. Of course, this position gains substance only by proposing specific forms of interaction. One might think that various such forms are possible. However, belief and probability do not mesh easily; the first difficulty is to find at all a workable coherent interactive scheme. Not many have seriously tackled this difficulty, and it is clearly Isaac Levi who has made the most elaborate proposal in this spirit. I return to it below.

[3] The analogous riddle is perhaps presented by Frank Jackson's color-blind Mary; but whether it can bar the reduction of the mental to the physical is highly contested.

[4] The first account that may be interpreted in this way is, as far as I know, the theory of belief functions of Shafer (1976). Probability measures are special belief functions, and consonant belief functions, as Shafer (1976, ch. 10) calls them, may be understood as representing belief (though I argued in Spohn (1990) that they do not do it adequately). Plausibility measures as elaborated by Halpern (2003) are to provide an overarching structure with many familiar theories as special cases. My probabilified ranking functions (see the appendix) provide yet another option for such unification.

Which other form might dualism take, if not interactionism?[5] *Separatism*, as I tentatively call it; what comes next to it in the philosophy of mind is perhaps psychophysical parallelism. Separatism is the view that acknowledges a useful theory of subjective probability and a useful theory of belief, and keeps them as coexisting but separate enterprises not in need of unification in an integral picture. In a way, this view draws the most negative conclusion from the lottery paradox by accepting it as unsolvable. Is this a feasible position? Yes, I hope so. At least, separatism is the label that best fits my own position. What might be its justification? I discuss this below.

This classification of views is useful as far as it goes. In particular, it emphasizes the importance of the lottery paradox, which forces one to take a stance within the spectrum of possibilities opened by it. Without such a stance one's epistemology would be incomplete. On the other hand, the classification is not entirely reliable. We need to look at the epistemological positions themselves; then we shall see that the dialectical situation is more complex than the classification reveals. So, let us inspect Levi's stance a bit more closely.

2. LEVI'S PICTURE

The first cornerstone of Levi's epistemology[6] is his notion of *full belief*. Full beliefs are free of any doubt; this is Levi's Peircean heritage. They are not in need of justification; they rather form the current base on which to proceed. However, full beliefs can change, the changes need be justified, and Levi is amply occupied with this justification.

Before turning to it, we have to locate probability. Full beliefs are undoubted or, as Levi often says, maximally and equally certain or infallible (from their own point of view), though corrigible (because they can change). They exclude certain possibilities and leave open others. They thus provide a standard of possibility, *serious possibility*, as Levi calls it. Serious possibilities define the space of inquiry, which at the same time is the *space of probability*; only serious possibilities (and sets of them) genuinely carry probability. Hence, full beliefs have probability 1, but not in the sense that I would bet my life on them, but by default, as it were, since full beliefs define the frame of probabilistic judgments. The possibilities excluded by them do

[5] Not epiphenomenalism; I have no idea what its epistemological analogy might be.

[6] Levi's epistemology is surprisingly consistent over the decades. Many essential elements are laid out already in Levi (1967) and still found today. The picture has become much richer over the years, and now and then an error (in his view) had to be corrected. I mainly refer to Levi (2004), which, among other things, beautifully summarizes Levi's view and its development.

not have any genuine probabilistic role since they are tied to probability 0. This is, basically, Levi's dualism in the above sense.

To see how the dualism develops into an interactionism, we have to look at Levi's very detailed ideas about how to change full beliefs. There are two basic movements, expansion and contraction of full beliefs. Levi provides detailed justificatory accounts for them; all other changes can only be indirectly justified by getting decomposed into such basic moves. Such decomposition is also required because the two basic moves are guided by entirely different principles. Let us look at them separately.

Expansion of full belief is an epistemic decision problem, according to Levi. On the one hand, one seeks valuable information; on the other hand, one wants to avoid error. Thus, one's acceptance is torn between stronger and weaker propositions. This decision problem is solved by determining the *expected epistemic utility* $EV^*(h)$ of each proposition h.[7] The simplest expression Levi derives for EV^* (cf. Levi 2004, ch. 3) is

$$EV^*(h) = Q(h) - q \cdot M(h).$$

Here, $Q(h)$ is the subject's *credal probability* for h. M is the subject's *informational value determining function* (the *undamped* version, to be precise), which Levi argues to behave like mathematical probability, although its interpretation must be sharply distinguished from that of credal probability. h's informational value itself is then given by $1 - M(h)$; thus, the stronger a proposition, the larger its informational value. q, finally, is an *index of boldness* between 0 and 1.

This label finds its explanation in Levi's *decision rule*. He argues that one should reject all and only those strongest propositions (or atoms of the propositional algebra under consideration) not yet excluded by one's full beliefs that carry negative expected epistemic utility. Equivalently, one should accept precisely the negations of these propositions (and their logical consequences). Hence, the greater one's q, the bolder one is in rejecting (and accepting) propositions.

Now it is evident why I call Levi's view interactionistic; on the one hand, full beliefs delimit a space for subjective (credal) probabilities; on the other hand, credal probabilities are crucial for expanding full beliefs. In this way, belief and probability depend on each other, though none gets absorbed by the other.

[7] The talk of propositions is mine. Levi prefers to talk of conjectures, hypotheses, etc. Ultimately, he analyzes what doxastic attitudes refer to in terms of sentences and sets of sentences. This is an aspect I cannot, and need not, lay out here.

Another important notion finds its place in the picture sketched so far. Instead of fixing the index of boldness in advance, one may as well consider the maximum index at which the proposition h is still unrejected and define it as its *degree of unrejectability* $q(h)$ (which is 1 if h is not rejected for any q). The more easily a proposition is rejected, the more surprising it is to learn that it obtains. Hence, the *degree of potential surprise* of a proposition h may be defined as $d(h) = 1 - q(h)$. The dual notion is the *degree of belief* or *plausibility* of h, which is defined as $b(h) = d(\text{non-}h)$. That is, h is the more plausible, the more surprising its negation.

These definitions entail: All and only propositions excluded by full beliefs are maximally surprising. All and only full beliefs themselves are maximally plausible. If h is surprising (or plausible) to some positive degree, non-h is not; that is, if $d(h) > 0$, then $d(\text{non-}h) = 0$, and if $b(h) > 0$, then $b(\text{non-}h) = 0$. The potential surprise value of a disjunction is the minimum of the surprise values of its disjuncts; that is, $d(g \text{ or } h) = \min(d(g), d(h))$. And the degree of plausibility of a conjunction is the minimum of the plausibility degrees of its conjuncts; that is, $b(g \text{ and } h) = \min(b(g), b(h))$. Thus, as Levi emphasizes, the degrees of potential surprise and the dual degrees of belief are constructed so as to satisfy precisely the axioms already proposed by Shackle (1949).

There is no saying whether credal probabilities or degrees of plausibility have more claim on being called degrees of belief. They coexist, they are different, but they are related (via the presented chain of definitions). Credal probabilities are relevant for assessing risk and for determining expected values (or utilities) in practical and epistemic decision making. By contrast, according to Shackle's degrees of belief, the set of propositions believed at least to some degree $x > 0$ is consistent and deductively closed, just like the set of full beliefs; hence, they are suited for assessing changes of full beliefs through expansion. This must suffice as a very brief description of Levi's account of expansion.

Contraction is also an epistemic decision problem, but quite different from expansion. There is no unfailing truth guarantee for our full beliefs; they must be conceived as corrigible, and indeed sometimes we find reasons to give up some of them. The problem is, how? In contrast to expansion, this problem is one-dimensional, as it were; there is no risk of error in contraction, since one does not acquire any beliefs at all in contraction and a fortiori not any false beliefs. Therefore, the only parameter guiding contraction is informational value, and since one is losing information in contraction, this loss must be minimized.

Thus put, contraction appears simpler than expansion. There is, however, a severe complication. At first, one may think one can apply the same informational value determining function M that was used in expansion. However, if contraction minimizes loss of information in this sense, the resulting contraction behavior is heavily defective, from an intuitive as well as a theoretical point of view. This is the starting point of a quite involved discussion in which Levi arrives at the result that loss of information must be measured not by the undamped version of M mentioned above which behaves like mathematical probability, but rather by a *damped* version. Of course, everything then depends on the specific damping applied – for all this see Levi (2004, ch. 4).

I cannot go into details here. But there is a tricky point of interpretation that I have to explain briefly. Damped informational value also satisfies the above-mentioned Shackle axioms. Therefore, one must take great care not to confuse the various uses of this structure. Damped informational value is not degree of plausibility, simply because plausibility, like probability, relates only to serious possibilities, whereas damped informational value refers to the possibilities excluded by full belief in order to govern contraction. One might be tempted then to interpret damped informational values as prior degrees of plausibility, that is, as plausibility degrees one had in a (hypothetical) state before acquiring any substantial full beliefs. But this does not fit, either. Prior plausibility governs expansion of that (hypothetical) ignorant state by bringing informational value and risk of error into a balance, but this explanation does not apply to damped informational value.

AGM belief revision theory was able to explain contraction (and revision) by an underlying entrenchment ordering, as it was called (cf. Gärdenfors 1988, ch. 4). Again, though, Levi discovers subtle differences. He proposes to reconstruct entrenchment in terms of damped informational value. But they are not the same, and hence differences in the view of contraction emerge. (For all this, see Levi 2004, ch. 5.)

To summarize: According to Levi, it must be informational value that guides contraction, and it must be a specific damped version in order for contraction to work properly. Other approaches to contraction are argued to be misguided, even if they have the same or a similar formal structure.

Of course, there are other changes of full beliefs. In particular, there are revisions in which one is forced to accept a hitherto unexpected or rejected proposition h, and residual shifts, as Levis calls them, where one replaces a formerly accepted proposition f by another proposition g. Any such change can be decomposed into a contraction and an expansion. Levi, however, insists that it *must* be so decomposed, since this is the only way to rationalize it. As explained, contractions and expansions have their own but quite different justifications that can be applied successively, but not mixed to yield direct justifications of other forms of changes.

So much for Isaac Levi's epistemological picture. The richness of details of his many books could not even faintly be displayed here, but I hope I have fairly represented his basic moves. My picture is surprisingly different, so much so, indeed, that it would be hopeless to start discussing details. Hence, I am going to describe my picture at the same basic and informal level. If you want to be fair to me, however, please simply forget Levi's picture for the next section. Any attempt at translation would be premature and misleading. The two accounts are brought together only in the final section.

3. MY VIEW

Like Levi I accept the moral of the lottery paradox, and like Levi I cannot simply discard the notion of belief. Thus I am a dualist, too. However, I start as a methodological separatist. If the first aim, a reduction of belief to probability, is not feasible, then my next aim is to develop an account of belief as well and as far as possible – separately, because the clarification of the relation of that account to probability can only come afterward as a third step. Whether I end up as a genuine separatist is not so clear. I would like to see myself as at least formally returning to reductionism (see the appendix); this would be theoretically most satisfying. Soberly, though, I prefer to present myself as a separatist (see below).

In developing an account of belief, I would very much like to naïvely talk of belief *simpliciter*. However, people talk of full belief, strong belief, weak belief, and so on, and each adjective is backed up by a whole theory. To distinguish myself from all this, I introduced the notion of *plain belief*, hoping that "plain" would not be conceived as a qualification at all. Twenty years later I feel that the qualification is misleading rather than helpful. Therefore I return here to talk of belief *simpliciter*. The main reason is that I cannot find that belief is in any way ambiguous. Belief is vague. When asked "*p*?" I usually say: "Yes" or "No" or "I don't know," in which cases I clearly express my belief or my ignorance. Quite often, I also start qualifying: "Yes, I am absolutely sure" or "Yes, I think so" or "I guess so" or "Presumably," and so on. We have here a rich vocabulary, indicating first that belief admits of degrees (which need not be conceived as probabilities) and second that these degrees open a range of vagueness within which it is hard determinately to call someone a believer or a nonbeliever. But this does not make belief ambiguous; ambiguity, it seems to me, is imported only by the various theories about it. Likewise, I reject the purely theory-induced distinction between "belief" and "acceptance."

As everybody, I start analyzing belief *simpliciter* as a set of propositions (or sentences) believed true (by a certain person at a certain time). And in order to start theorizing at all, I, like most, take such a *belief set* to be consistent and deductively closed. I like Levi's defense of these assumptions in terms of commitment (cf. Levi 2004, sec. 1.2).

So far, the analysis is purely static. To be complete, an account of belief, as of any phenomenon in time, must be dynamic. Clearly, the functioning and malfunctioning of memory is the most important determinant of the dynamics of belief. But presupposing perfect memory and excluding other arational influences, it is experience, perception, observation that drives the dynamics of belief and that occupies the interest of epistemologists such as me. Hence, my focus is on belief revision, which may be a simple expansion or a genuine revision, depending on the compatibility of experience with expectations.

At first, I thought that AGM belief revision theory (cf. Gärdenfors 1988, ch. 3) offers a perfect account of belief revision, but soon I realized that it is unsatisfactory because it violates what was later called the principle of categorical matching (cf. Gärdenfors and Rott 1995, p. 37) and hence offers only a restricted account of iterated belief revision.

Thus, I came (Spohn 1983, 1988) to introduce ranking functions (a terminology proposed by Judea Pearl and Moises Goldszmidt to which I happily converted later on). Given a set W of all possibilities under consideration, a *ranking function* κ for W is a function from the power set $\mathcal{P}(W)$, the set of propositions, into the extended set of natural numbers $\mathbf{N} \cup \{\infty\}$ such that $\kappa(W) = 0$, $\kappa(\emptyset) = \infty$ and for all propositions A, $B \kappa(A \text{ or } B) = \min\{\kappa(A), \kappa(B)\}$. My standard explanation (thanks to Isaac Levi) is that ranks are *degrees of disbelief*: $\kappa(A) = 0$ says that A is not disbelieved at all, $\kappa(A) = n (n \geq 1)$ says that A is disbelieved to degree n. Hence A is *believed* relative to κ iff $\kappa(\text{non-}A) > 0$ (but see my important qualification in the next section). There is no difficulty in introducing the dual notion of degrees of positive belief, but ranks of disbelief are preferable for various reasons.

So far, the formal structure has many predecessors and Levi is right to point out that Shackle (1949) was the first: The above functions d of potential surprise and my ranking functions κ are obviously governed by the same axioms (the difference in the range of these functions, which is not arbitrary, need not concern us here). Of course, the informal predecessors reach back much further. However, the reference to predecessors is also misleading – in particular it has misled Levi, I think – because it imports all the old and perhaps unwanted interpretations and because it obstructs the view on variations and new developments.

Indeed, what is doing all the work in my account of belief is the definition of conditional ranks. This is a topic that received remarkably little attention by the predecessors.[8] The only explanation I have is that the dynamic perspective so central for me became focal within philosophy only in the 1970s and that it is only in this perspective that conditional ranks unfold their power and beauty.

To be a bit more precise, the *conditional rank* of B *given* A is defined as $\kappa(B \mid A) = \kappa(A \text{ and } B) - \kappa(A)$, if $\kappa(A) < \infty$; otherwise, it is undefined. In other words, the degree of disbelief in A and the degree of disbelief in B given A add up to the degree of disbelief in A-and-B. This is intuitively convincing; it suggests a Cox-like justification of the ranking structure; it opens a way to measuring ranks (on a proportional scale) via contraction behavior (still to be explained); and it provides the clearest justification of the crucial last axiom for ranking functions: Given the definition of conditional ranks, this axiom is equivalent to either $\kappa(A \mid A \text{ or } B) = 0$ or $\kappa(B \mid A \text{ or } B) = 0$ (or both), and this is simply a conditional consistency requirement saying that A and B cannot both be disbelieved given A-or-B.

In thorough-going analogy to probability theory, conditional ranks allow a substantial development of ranking theory. Qualitative confirmation theory did not get off of the ground in the 1950s and 1960s, but now we can make the basic idea of confirmation theory work. A *confirms* B, or A is a *reason* for B (as I prefer to say in order to make ranking theory interesting for the traditional epistemologist), iff A is positively relevant to B, that is, iff $\kappa(\text{non-}B \mid A) > \kappa(\text{non-}B \mid \text{non-}A)$ or $\kappa(B \mid A) < \kappa(B \mid \text{non-}A)$. We can define (conditional) dependence and independence with respect to ranking functions. Indeed, we can develop a full-blown ranking analogue to the theory of Bayesian nets, which is a beautiful formal representation of our intuitive ways of thinking, as such still grossly underrated by philosophers.

On this basis, then, my picture of belief dynamics is as follows: The subject directly perceives that A and must hence revise his or her body of beliefs so as to contain A. But no, this is wrong from the start; the idea of direct perception and its false understandings have done a lot of damage to philosophy in the last 350 years. I better begin in this way: Experience affects the subject's beliefs; somehow, among all the propositions under consideration, it is the proposition A that is affected first and thus believed to some degree n (i.e., $\kappa'(\text{non-}A) = n$, where κ' is the posterior ranking function). This degree n is part of the experiental process, and it can vary. I read in the newspaper that A,

[8] Shackle is an exception; conditional potential surprise played an important role in his theory (cf. Shackle 1969, part IV). There (p. 205) he even briefly considers the definition of conditional ranks I use, but he rejects it. He prefers the postulate $d(A \text{ and } B) = \max\{d(A), d(B \mid A)\}$, which in fact takes $d(B \mid A)$ as an undefined primitive and which he takes to be simpler and less unrealistic.

I hear with my own ears that A, my wife tells me that A, I see that A: These are four ways to learn A with increasing firmness of belief (since I trust my wife more than my ears). So, A as well as n form the input of the doxastic change that I take as given. But how should the subject respond to this experience? With Jeffrey (1965, ch. 11) I say that the prior ranks conditional on A and conditional on non-A are not changed through this experience. That is, if κ is the prior ranking, we have for any proposition B:

$$\begin{aligned}\kappa'(B) &= \min\{\kappa'(A \text{ and } B), \kappa'(\text{non-}A \text{ and } B)\} \\ &= \min\{\kappa'(A) + \kappa'(B \mid A), \kappa'(\text{non-}A) + \kappa'(B \mid \text{non-}A)\} \\ &= \min\{\kappa(B \mid A), n + \kappa(B \mid \text{non-}A)\},\end{aligned}$$

where the first steps are transformations according to the ranking calculus and the last step realizes the crucial assumptions just explained. This κ' is what I describe as the *A,n-conditionalization* of κ (cf. Spohn 1983, sec. 5.3, and 1988, sec. 5). And it has many special cases. In particular, it is an expansion in case $\kappa(A) = \kappa(\text{non-}A) = 0$ and $n > 0$; it is a genuine revision in case $\kappa(A) > 0$ and $n > 0$; and it is a genuine contraction in case $\kappa(\text{non-}A) > 0$ and $n = 0$. This rule of belief change can be iterated indefinitely as long as the prior rank of the experiental proposition is finite (i.e., $\kappa(A)$, $\kappa(\text{non-}A) < \infty$); this is why I usually assumed that $\kappa(A) = \infty$ only for $A = \emptyset$. To this extent, I may claim to have offered a complete dynamics of belief. This must suffice as a sketch of my account.

Can this picture be combined with subjective probability theory? Yes and no. Ranking theory is very similar to probability theory; there is indeed an algorithm for translating probabilistic into ranking theorems that is almost guaranteed to work. Therefore, it is perfectly natural, and not a gerrymander, to integrate both into one notion that I call probabilified ranking functions (or ranked probability measures) – see the appendix for the definition. All the theoretical developments I sketched above work equally well for that notion. Hence, we may speak here of a genuine theoretical unification and of a reduction of both probability and belief to that notion. One may also call this simply an extended probabilistic point of view. In any case, if one accepts this unification, one returns to reductionism.

I do not dare do so. For the lottery paradox raises its ugly head again. Unavoidably, this unification has the consequence that disbelief in A, that is, $\kappa(A) > 0$, entails $p(A) = 0$. In other words, believing in A, even to the lowest degree, means giving probability 1 to A. This strikes me as counterintuitive. Believing in A is a very ordinary affair; it is not being probabilistically maximally certain of A, betting everything I have on A. Hence, from the point of view of the intended interpretation, the two notions, ranks and probabilities, *do not mesh*, as far as I can see. This is why I feel the unification to be artificial despite its theoretical beauty and why I prefer to develop the two theories separately, though always with an eye on the format parallel. Thus I remain a separatist (and may perhaps even be called a parallelist). But I would be glad if someone teaches me better.

4. COMPARISONS

These sketches suffice to make clear two things. First, that Isaac Levi and I are dealing with the same issues in the same spirit; that is why a comparison is relatively straightforward. Second, that our conceptions differ widely; indeed, we diverge on nearly every point. Let me collect here the differences, and let me try to identify their deeper sources. We shall find a divergence in basic attitudes, where it becomes nearly impossible to tell who is right and who is wrong.

Some differences have already been fairly explicit: (1) I have explained why Levi takes expansion and contraction to be the basic moves in belief change, and I have explained why I start considering revision that led me to state the role of *A*,*n*-conditionalization that in turn embraces contraction as a special case. (2) My contraction then endorses the AGM orthodoxy on contraction; this was my intention. Hence, I also endorse the recovery postulate for contraction not yet mentioned so far (cf. Gärdenfors 1988, p. 62). By contrast, Levi rejects the recovery postulate and he takes great pains to construct contraction accordingly; indeed, the whole chapter 4 of Levi (2004) is devoted to this issue. (3) I mentioned the importance of the principle of categorical matching for my line of reasoning. Levi (2004, sec. 5.11), however, does not accept it as a regulative principle for theories of belief change. (4) This difference is explained by a different approach to iterated belief change. I have sketched how this problem drove the development of my theory. Levi does not see such a big problem here. For him, informational value is the only parameter governing contraction and an essential parameter governing expansion that he assumes to be relatively stable, at least within a given inquiry. Thus, there is no need to develop a dynamic account for it. Given this stable parameter, Levi can, of course, also explain iterated changes.

One may certainly discover more differences of this specificity. This is good; if theories do not get down to such details, they are of no worth. In principle, we should carry out such differences. However, I refrain from starting an argument here. The specific differences only reflect deeper strategic ones, and without settling the latter, agreement on the former is spurious.

(5) The first strategic difference is – indeed, this was the red line of this chapter – that Levi and I go together for a while as dualists and then split as interactionist and separatist.

(6) This difference is entailed, it seems to me, by a deep difference over the notion of belief. Levi's notion of full belief and the accompanying combination of infallibility and corrigibility do not fit into my picture, and my notion of belief and my rejection of this combination do not fit into his. Here, our views are entirely at cross-purposes. In a way, this conclusion is reached by Levi (2004, secs. 1.6 and 3.5), too.

Levi there suggests that I am a Parmenidean skeptic. This is a person, according to Levi, for whom the only standard of serious possibility is logical or conceptual possibility and for whom full beliefs are restricted to unrevisably or incorrigibly a priori beliefs. Such a person is called a skeptic because for Levi the full beliefs of a subject precisely set the undoubtful frame within which he or she is conducting his or her inquiries. Thus, if the Parmenidean has such a narrow notion of full belief, he or she has a very large notion of doubt.

According to this suggestion, my beliefs can then be understood as expansions of Parmenidean full beliefs. This is so because Levi has assigned a firm place to Shackle's functions of potential surprise in his epistemological picture as determining which beliefs to accept according to the chosen index of boldness and because my ranking functions work the same and thus apparently do the same as the functions of potential surprise. Indeed, I am a particularly bold expander according to Levi, since by calling the minimal rank 1 already disbelief I am in effect applying the maximal index of boldness.

Maybe this is the relatively best embedding of my views into Levi's. Still, instead of forcing this embedding one better acknowledges that it does not fit at all. I grant that I have the notion of unrevisably a priori beliefs. Their negations are disbelieved with maximal rank ∞ for which no dynamics is explained; this is why they are unrevisable. Unlike Levi, though, I do not see this notion as a remnant of a degenerating research program; on the contrary, the research program on apriority puts forth blossoms again (cf. Spohn 2000).

Be this as it may, it is not correct to interpret my defense of a priori beliefs as holding a particularly narrow and rigid notion of full beliefs. I have no full beliefs at all. I have beliefs; I need not justify them unless questioned, but of course, I can justify them even if unquestioned. That is, each proposition receives that rank that is assigned to it by my balance of reasons for and against it (in the sense explicated above), and if this balance of reasons assigns a positive belief value to it (i.e, a positive rank of disbelief to its negation), then I believe it and do not doubt it. I do not see why I should treat my beliefs as doubtful, unless they are full. But, of course, my beliefs are revisable, just as Levi's full beliefs are.

Moreover, I need not be attached to being maximally bold (in Levi's terms). As Matthias Hild has pointed out to me in conversation, I am not forced to say that A is believed relative to κ iff $\kappa(\text{non-}A) > 0$. I could as well define stricter notions of believing A, say as $\kappa(\text{non-}A) > n$ for some $n > 0$. The logic of belief always comes out the same. We might indeed say that the logic of belief is represented by the ranking structure, however we precisely map the vague notion of belief onto this structure.

Here is a difference, I think, we should not try to bridge. At least it explains our attitudes toward probability. With his standard of full belief and serious possibility Levi can delimit a realm of probability and thus develop an interactionist picture. Without such a standard I can only keep belief and probability separate, or probabilify each level of belief (as explained in the appendix).

(7) The differences go still deeper. It seems to me that even our methodologies are opposite. Levi approaches epistemology in a constructive-synthetic way, whereas I rather have an analytic-structural view. Levi has enormously detailed stories to tell about our epistemological wheelwork; this is, of course, one of the fascinating aspects of his work. There are small wheels and big wheels, and Levi meticulously explains how they interact in order to produce judgment and belief. A particularly big wheel is informational value, but it is driven by several smaller wheels, such as explanatory power, simplicity, generality, and specific interests.

I am not telling any such stories; I even refuse to do so, since I am not so sure of all these wheels. What is simplicity? What is explanatory power? What is overall informational value? And so forth. What is the measurement theory for all these magnitudes? None, I suppose. As long as such questions are not well answered, I remain in doubt as to whether the wheelwork is a real, hypothetical, or imaginary one. This is why I prefer to tell my structural story. Whatever the inner epistemological workings, the resulting structure must be such and such, viz. a ranking structure, if my arguments and in particular my measurement theory for ranks hold good. My way of proceeding is then just the reverse of Levi's. Starting with this structure, my hope is to be able to analyze and lay bare some of these inner workings: lawhood, explanation, causal inference, the truth-conduciveness of reasons, and so on.

Clearly, both the analytic and the synthetic method are legitimate. Presently, I do not see a clear meeting point, but maybe the methods can be joined. In the end, overall success decides; until then both projects have to be pushed ahead as far as possible.

(8) This methodological point helps to explain why Isaac Levi and I have such difficulties in agreeing on the relation between Shackle's functions of potential surprise and my ranking functions. As I have explained, Levi has firmly located Shackle's functions within his account of expansion and tends to put ranking functions into the same drawer. Damped informational value governing contraction may have the same formal structure, but has an entirely different use and interpretation. From Levi's constructive point of view this is how one must see the situation.

From my structural point of view I do not care whether expected epistemic utility or damped informational value generate the relevant structure. Since it is the same structure, I comprise it in my ranking functions and apply it across the board. Hence, I insist that ranking functions govern both expansion and contraction in a unified way (indeed, unified by A,n-conditionalization), even if this must appear as mixing apples and oranges to Levi.

(9) I see a final difference, perhaps the deepest of all, which motivates the issue about methodology. It is a difference about how to conceive of justification in epistemology. In a way, Levi is very modest; he wants to justify not current beliefs, but only changes of belief. Concerning the latter, he is less modest. As he repeatedly says, his inquiry of the epistemological wheelwork is to uncover in detail the justificatory structure supporting belief change.

I do not say that Levi is pursuing an objectivistic notion of justification here. He emphasizes the context-dependence of all his parameters. His index of boldness is entirely subjective; there is no prescription of a wise choice. Informational value is basically subjective; there is no need for people to agree on it. Still, I sense the remnants of an objectivistic notion in Levi's work. Even if a lot of wheels are subjective, his hope is to identify at least some objectively valid components (such as explanatory power) and at least some objectively valid connections between the components (such as expansion according to expected epistemic value).

This is not my picture. I am immodest in thinking that I can also justify my current beliefs (see (6) above), but in doing so I apply a more radically subjectivistic notion of justification. Like Levi, I take it as axiomatic for rational subjects that it is reasons that drive their doxastic dynamics. But I do not pretend to have an antecedent notion of reasons that I could plug in here. Rather, I read the axiom inversely – this is why I am also reversing the methodological order – as saying that for rational subjects reasons are whatever drives their doxastic dynamics. Look at their (actual and potential) dynamics, and you know what their justifications are. Perhaps Levi is right in calling me a skeptic.

On the other hand, the difference is not so big as it may seem. Of course, I am also discontent with an entirely subjectivistic point of view, and then I am starting with the inquiry into what I call the a priori structure of the space of reasons (see Spohn 2000). Still, I very much doubt that Levi's and my inquiry can be brought to converge.

Facing differences so profound, I find it hard to start an argument; any argument is bound to end in a draw. Hence, the point of this chapter was only to explain and to clarify the dialectic situation between Isaac Levi and me for the benefit of future discussion.

APPENDIX

Let me briefly explain what a probabilified ranking or a ranked probability is by working up from standard probability. Let W be a set of possibilities, which we assume to be finite for the sake of simplicity. Any subset of W is a proposition.

Then, p is a *probability measure* for W iff p is a function from $\mathcal{P}$, the set of propositions, into $\mathbf{R}$ ($=$ the set of real numbers) such that for all $A, B \in \mathcal{P}$:

(1) $p(A) \geq 0$,
(2) $p(W) = 1$,
(3) if $A \cap B = \emptyset$, then $p(A \cup B) = p(A) + p(B)$.

If $p(A) \neq 0$, the *conditional probability* of B *given* A is defined as $p(B \mid A) = p(A \cap B)/p(A)$.

There is a more general notion that takes conditional probability as basic. p is a *Popper measure* for W iff there is a set $\mathcal{C} \subseteq \mathcal{P}$ (the set of conditions) such that p is a function from $\mathcal{P} \times \mathcal{C}$ into $\mathbf{R}$ such that for all $A, B, C \in \mathcal{P}$:

(4) for each $C \in \mathcal{C}$ $p(. \mid C)$ is a probability measure for W,
(5) if $B \cap C \in \mathcal{C}$, then $p(A \cap B \mid C) = p(A \mid B \cap C) \cdot p(B \mid C)$,
(6) if $B \in \mathcal{C}$ and $p(A \mid B) > 0$, then $A \in \mathcal{C}$.

It is well-known (cf., e.g., Spohn 1986) that Popper measures and standard probability measures relate in the following way: $(p_0, \ldots, p_n)$ is a *sequence of probability measures* for W iff:

(7) each $p_i (i = 0, \ldots, n)$ is a probability measure for W,
(8) for each $j = 0, \ldots, n$ there is a C_j such that $p_j(C_j) = 1$ and $p_i(C_j) = 0$ for all $i < j$.

Such sequences and Popper measures are strictly equivalent in the following sense: For each Popper measure p for W there is exactly one sequence $(p_0, \ldots, p_n)$ of probability measures for W, and vice versa, such that for each $A \in \mathcal{P}$ and $C \in \mathcal{C}$:

(9) if $j = \min\{i \mid p_i(C) > 0\}$, then $p(A \mid C) = p_j(A \mid C)$.

Ranked probability measures, which are my proposal for unifying belief and probability, provide a further generalization. ρ is a *ranked probability measure* for W iff there is a function κ from $\mathcal{P}$ into $\mathbf{N} \cup \{\infty\}$ and a function π from $\mathcal{P}$ into $\mathbf{R}$ such that for all $A, B \in \mathcal{P}$:

(10) $\rho(A) = \langle \kappa(A), \pi(A) \rangle$,
(11) $\kappa(\emptyset) = \infty$ and $\pi(\emptyset) = 0$,
(12) $\kappa(W) = 0$ and $\pi(W) = 1$,
(13) $\kappa(A) < \infty$ iff $\pi(A) > 0$,
(14) if $\pi(A) > 0$, then there is exactly one C such that $\kappa(C) = \kappa(A)$ and $\pi(C) = 1$,
(15) if $A \cap B = \emptyset$, then $\kappa(A \cup B) = \min\{\kappa(A), \kappa(B)\}$ and

$$\pi(A \cup B) = \left\{ \begin{array}{l} \pi(A), \text{ if } \kappa(A) < \kappa(B) \\ \pi(B), \text{ if } \kappa(A) > \kappa(B) \\ \pi(A) + \pi(B), \text{ if } \kappa(A) = \kappa(B) \end{array} \right\}.$$

If $\kappa(A) < \infty$, the *conditional ranked probability* of B *given* A is defined as $\rho(B \mid A) = \langle \kappa(B \mid A), \pi(B \mid A) \rangle$, where $\kappa(B \mid A) = \kappa(A \cap B) - \kappa(A)$ and, if $\kappa(A \cap B) < \infty$ and C is the proposition such that $\kappa(C) = \kappa(A \cap B)$ and $\pi(C) = 1$, $\pi(B \mid A) = \pi(A \cap B \cap C)/\pi(A \cap C)$.

It is easy to see that such a ranked probability measure ρ is equivalent to a *ranked* sequence of probability measures that is like an above sequence satisfying (7) and (8) with the only difference that the sequence is not indexed by $\{0, \ldots, n\}$, but by the finite ranks occupied by ρ, that is, by $\{\kappa(A) \mid A \in \mathcal{P}, \kappa(A) < \infty\}$.

Ranked probability measures look like a hybrid, but, as emphasized, they show a unified behavior. One may conjecture the unified behavior is due to the fact that ranked probability measures are similar to nonstandard probability measures, since ranks may be conceived as orders of magnitudes relative to some infinitesimal. This conjecture is certainly true, but the precise extent of the similarity still awaits clarification.

Gärdenfors, Peter. 1988. *Knowledge in Flux: Modeling the Dyamics of Epistemic States.* Cambridge, Mass.: MIT Press.

Gärdenfors, Peter, and Hans Rott. 1995. "Belief Revision." In D. M. Gabbay, C. J. Hogger, and J. A. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming,* vol. 4: *Epistemic and Temporal Reasoning*, pp. 35–132. Oxford: Oxford University Press.

Halpern, Joseph Y. 2003. *Reasoning about Uncertainty*. Cambridge, Mass.: MIT Press.

Hintikka, Jaakko. 1962. *Knowledge and Belief.* Ithaca, N.Y.: Cornell University Press.

Jeffrey, Richard C. 1965. *The Logic of Decision*. Chicago: University of Chicago Press.

Jeffrey, Richard C. 1992. *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.

Kyburg, Henry E., Jr. 1961. *Probability and the Logic of Rational Belief.* Middletown, Conn.: Wesleyan University Press.

Levi, Isaac. 1967. *Gambling with Truth*. New York: Knopf.

Levi, Isaac. 2004. *Mild Contraction: Evaluating Loss of Information Due to Loss of Belief.* Oxford: Oxford University Press.

Shackle, George L. S. 1949. *Expectation in Economics*. Cambridge: Cambridge University Press.

Shackle, George L. S. 1969. *Decision, Order, and Time in Human Affairs*, 2nd ed. Cambridge: Cambridge University Press.

Shafer, Glenn. 1976. *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

Spohn, Wolfgang. 1983. *Eine Theorie der Kausalität*, unpublished Habilitationsschrift, Munich.

Spohn, Wolfgang. 1986. "The Representation of Popper Measures." *Topoi* 5: 69–74.

Spohn, Wolfgang. 1988. "Ordinal Conditional Functions: A Dynamic Theory of Epistemic States." In W. L. Harper and B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, pp. 105–34. Dordrecht: Kluwer.

Spohn, Wolfgang. 1990. "A General Non-Probabilistic Theory of Inductive Reasoning." In R. D. Shachter, T. S. Levitt, J. Lemmer, and L. N. Kanal (eds.), *Uncertainty in Artificial Intelligence* 4, pp. 149–58. Amsterdam: Elsevier.

Spohn, Wolfgang. 2000. "Über die Struktur theoretischer Gründe." In J. Mittelstraß (ed.), *Die Zukunft des Wissens. Akten des 18. Deutschen Kongresses für Philosophie*, pp. 163–76. Berlin: Akademie Verlag.

10

Isaac Levi on Abduction

Maurice Pagnucco

> Progress is impossible without change; and those who cannot change their minds cannot change anything.
>
> *George Bernard Shaw*

The purpose of this chapter is to survey Isaac Levi's conception of abduction and to contrast it with other work on abduction to be found in the literature. In particular, I concentrate on developments in abductive logics within the area of artificial intelligence where the main concern is to implement an abductive form of inference. I contrast the views to be found there with Levi's conception of abduction. In large part, I show that Levi's notion of abduction is quite distinct to what is found in the artificial intelligence literature.

Deduction, induction, and abduction have variously been viewed as essential elements in the stages of scientific inquiry. Isaac Levi's work has gone a long way toward clarifying these notions and identifying their roles in a reasoner's process of inquiry. In brief, abduction is used to identify potential answers to an inquiry; induction selects the most desirable of these given the reasoner's requirements; and deduction determines the consequences of this selection. I seek to examine abduction's role in further detail by looking at Levi's notion of this process and contrasting it against other views. While I have less to say about induction and deduction, I also touch on them as the three processes form a coherent whole in Levi's work.

I begin in the next section with a brief overview of Peirce's theory of abduction. In section 2, I survey Isaac Levi's ideas on abduction as they are applied to deliberate expansion. Section 3 looks at notions of abduction in artificial intelligence and contrasts them with Levi's work. In section 4, I discuss an alternative view of abduction, induction, and deduction and, again, compare it with Levi's work. I conclude with a summary in section 5.

One of the most urgent tasks of serious inquiry is the construction of informative potential answers to the question under study. This is the task of *abduction*, as C. S. Peirce called it. (Levi 1996, p. 161)

Charles Sanders Peirce introduced the term *abduction*. However, his ideas on abduction developed over time, undergoing a significant transformation throughout the course of their development. This section is based largely on the work of Fann (1970) and his treatment of the development of Peirce's conception of abduction.

In his earlier theories of inference, Peirce considered abduction, induction, and deduction to be three separate and fundamental forms of inference. He classified these inferential methods into two categories: *explicative inference*, where the conclusion follows necessarily from the premises, and *ampliative inference*, where it does not. He noted that from explicative inference to ampliative inference the "security" (certainty) of the inference decreases, while the "uberty" (productivity) increases (cf. Fann 1970, p. 8). At this early stage, Peirce's views can be summarized by the following logical forms (the example to the right of each figure is taken from Peirce (1931–58, vol. 2, p. 623).

DEDUCTION

$\forall x.P(x) \supset Q(X)$	All the beans from this bag are white
$P(a)$	These beans are from this bag
$Q(a)$	These beans are white

INDUCTION

$P(a)$	These beans are from this bag
$Q(a)$	These beans are white
$\forall x.P(x) \supset Q(X)$	All the beans from this bag are white

ABDUCTION

$\forall x.P(x) \supset Q(X)$	All the beans from this bag are white
$Q(a)$	These beans are white
$P(a)$	These beans are from this bag

We return to these when we discuss the nature of abduction as it is viewed in the artificial intelligence literature. In many cases, it is this earlier view of abduction that has been investigated there.

As his ideas developed, Peirce maintained that abduction, deduction, and induction constituted three stages of scientific inquiry (see Fann 1970, pp. 31–2):

1. abduction proposes hypotheses
2. deduction derives the consequences of the hypotheses
3. induction tests or verifies hypotheses.

Two major factors must be considered when abduction proposes hypotheses (Fann 1970, p. 41):

1. constructing or determining hypotheses
2. selecting the "best" or most plausible hypothesis from among these.

As we see in the next section, Levi's notion of abduction is preoccupied with the first of these. The second is the task of induction. Together, they constitute parts of the process of deliberate expansion that models the reasoner's task in attempting to acquire new information during inquiry.

Peirce suggests three major considerations in the selection of the best hypothesis (Fann 1970, p. 43):

1. a hypothesis must be capable of explaining the facts
2. a hypothesis must be verifiable – in particular, through experiment
3. considerations of "economy" should guide the choice of best hypothesis.

Since we are concerned with Levi's views on abduction and, as far as he is concerned, this does not concern itself with the selection of the best hypothesis – rather, the determination of which hypotheses to consider during inquiry – these considerations are, in a sense, secondary to our concerns here. However, they will be useful for when it comes to looking at the way abduction has been treated in artificial intelligence.

One interesting point about Peirce's ideas on abduction is that, with regard to the considerations for selecting the best hypothesis, truth plays no part in the selection. A hypothesis can be selected purely because it appears to be the most easily verifiable. If later it is determined to be false, an alternative hypothesis can be considered.

> As Peirce notes, the "conclusion" of an abduction can entail no error; for such a conclusion is the mere entertaining of a hypothesis for further test, scrutiny, and inquiry in an effort to answer some demand for information. (Levi 1980, p. 42)

Levi's conception of abduction largely coincides with Peirce's later views. His interests in abduction lie in an attempt to develop a theory of belief expansion where a reasoner is involved in adding to his or her belief corpus. Levi considers two forms of expansion: *routine expansion* and *deliberate* (inductive) *expansion*. Routine expansion is an instinctive form of expansion of a reasoner's belief corpus. It is not guided by any recourse to deliberation. As Levi puts it:

> [r]outine expansion is the result of implementing a program, to which the agent is committed beforehand, for using the testimony of the senses or of other agents as inputs where the outputs are expansions. (Levi 1984, p. 90)

In artificial intelligence this reactive behavior is identified with *tropistic* agents (Genesereth and Nilsson 1987). As the name suggests, this expansion behavior comes about through routine.

However, it is deliberate expansion that will occupy our attention here. Deliberate (or inferential or inductive) expansion is a reasoned form of expansion of the reasoner's belief corpus. To Levi, like Peirce, abduction is part of the process of deliberate expansion, as is induction (and, to a certain extent, deduction). In brief, it is that part of the process that occupies itself with formulating answers or explanations to inquiry. In particular, abduction is concerned with identifying "potential answers" or options that can settle the reasoner's inquiry. Discriminating among these potential answers and subsequently selecting one that is to be used in expanding the reasoner's belief corpus is another part of the process and one that is dealt with by induction. We shall have little to say about this part of the process here.

Levi views the abductive process of determining potential answers as an iterative one. That is to say, when the reasoner embarks on the process of deliberate expansion, he or she may have very little notion of what may constitute a potential answer or the range of options available to him or her. Hence, the process can begin with no potential answers (beyond maintaining the status quo or stumbling into inconsistency). On further refinement and contemplation, different potential answers may become apparent.

For Levi, the quest for potential answers is phrased with respect to an *ultimate partition*. In Levi's view:

> An ultimate partition for me represents the set of strongest consistent potential answers available to the agent at the given stage of inquiry and not the strongest consistent potential answers which in some obscurely defined sense are "in principle" available. (Levi 1984, p. 95 n. 1)

Using Levi's notation, a reasoner's belief corpus K is expressed in a (fixed) language L.[1] An ultimate partition U is relative to the reasoner's belief corpus K,[2] and as the name suggests, it is a partitioning of sentences in L. The following further restrictions apply (cf. Levi 1984, p. 93):

1. Each element of U is consistent with K; together these elements are exclusive and exhaustive relative to K
2. A potential answer relative to U and K adds sentences in the set $K_g - K$ to the reasoner's belief corpus K. The sentence g is true if and only if all the elements of K are true and an element of some subset of U is true. K_g represents the deductive consequences of K and the sentence g.

The sentence g represents a potential answer and "is a strongest sentence added to the evidence via induction" (Levi 1984, p. 93). This sentence is, in substance, equivalent to the set $K_g - K$. In this light we can see the role of the ultimate partition.

> The elements of X's ultimate partition at a given stage of inquiry consist of those hypotheses whose acceptance as strongest represent strongest consistent potential answers to X's demands for information and his current ability to identify potential answers gratifying demands for information. (Levi 1984, pp. 94–5)

As mentioned above, when the reasoner begins an inquiry, he or she may have no potential answers at hand. To Levi, at this point the ultimate partition consists of a single element that can be any part of the reasoner's belief corpus K. That is to say, accepting this as a potential answer would result in no change to K. As the process of inquiry proceeds, a number of potential answers may be identified.

Notice that the first restriction on ultimate partitions above stipulates that the elements of U are exhaustive relative to K. If the potential answers identified at any stage of the inquiry do not satisfy this restriction, then one can

[1] For my purposes, I am content to consider L to be finite. However, Levi considers languages that are not bound by this restriction.

[2] We omit the subscript (U_K) unless it is unclear to which belief corpus we are referring.

add a *residual hypothesis*, a potential answer stating that all other potential answers are false.

> X may seek a theory to explain phenomena in some domain and succeed in identifying T_1 and T_2. Relative to K, they are exclusive and consistent but not exhaustive. To form an ultimate partition, a residual hypothesis asserting the falsity of both T_1 and T_2 will be required.... This is so even though the residual hypothesis has little explanatory merit. (Levi 1984, p. 94)

Refinement of the ultimate partition may continue, through further partitioning of any of its elements, for instance, until the reasoner is satisfied. Once this refinement has ceased, the abductive process is over and induction is applied to determine a "best" answer from among all the potential answers. When, exactly, to end abduction is not specified in any formal way. This is up to the reasoner to judge. Levi does note that if induction were to lead to the residual hypothesis being chosen as the best answer, then this would be a timely moment to resort to abduction once more in an attempt to resolve satisfactorily the inquiry under consideration.

While it is not our primary concern here since in Levi's view it is the role of induction, we shall briefly consider his ideas on selecting from among the potential answers identified in the ultimate partition. For Levi, the reasoner's main objective during deliberate expansion is to acquire new information that is free from error. Of course, these two needs – gaining new information and avoiding error – are not independent of one another. They can be seen to work against each other as the reasoner's thirst for new information is tempered by his or her anxiety of stumbling into inconsistency and error.

The needs of gaining information and avoiding error are captured by two utility functions. $T(g, x)$ is the utility of selecting g with the view to avoiding error. $C(g)$ is the utility of selecting g with the view to obtaining new information. These are reminiscent of the Peircean notion of "security" and "uberty." The (overall) utility of selecting g in response to the reasoner's quest to acquire new, error-free information can be represented as the weighted sum $V(g, x) = \alpha T(g, x) + (1 - \alpha)C(g)$. The parameter α can take values in the interval [0.5, 1] so that error does not override correctness.

3. THEORIES OF ABDUCTION IN ARTIFICIAL INTELLIGENCE

> Charles Peirce, who did so much to emphasize the importance of abduction as a task of inquiry, conceived of induction as the task of eliminating potential answers from among those identified via abduction.... The conclusions of abductions are conjectures that are potential answers to questions. (Levi 1991, p. 77)

Abduction has a relatively long history of study within artificial intelligence's short lifespan.[3] In artificial intelligence, approaches to abduction can be divided into two categories: set-cover based and logic based.[4] The set-cover based approaches consist of sets of effects and sets of causes and a representation of the interrelationships between them. In determining an explanation, the main idea is to construct a set of causes whose related effects are a superset of all of those observed. Examples of this approach include Peng and Reggia's (1990) *parsimonious covering theory* and Allemang, Bylander, and Josephson's (1987) *hypothesis assembly*. Our focus here, however, is on the logic-based approaches as they are more relevant to Levi's work.

One of the earliest works in artificial intelligence, which explicitly refers to the notion of abduction, is Pople's (1973) attempt at mechanizing an abductive method of reasoning. His task is diagnosis. Pople's view of abduction coincides with Peirce's earlier view and the schema presented in section 1. Given a theory of the domain and starting from a set of observations, he utilizes the resolution proof procedure (Robinson 1965) to suggest hypotheses when a literal cannot be resolved away. This way of computing abduction leads to the following definition, which is the most common to be found in artificial intelligence (cf. Levesque's 1989 look at abduction at the knowledge level).

DEFINITION. An abduction for a set of formulae (observations) Φ with respect to a domain (or background) theory Γ is a set of formulae Ψ such that the following two conditions are satisfied:

1. $\Gamma \cup \Psi \vdash \Phi$
2. $\Gamma \cup \Psi \nvdash \bot$

That is to say, the abduction Ψ, when added to the background theory Γ, accounts for the observations Φ, and furthermore, the abduction is consistent with the background theory. As such, it is a consistent "explanation" of the observations. Thus far, this is keeping relatively faithful to Levi's notion of abduction. The abductions generated by the definition above are potential answers to a certain degree. They are not, however, the potential answers of Levi. There is no sense in which the definition above generates an ultimate partition, although elements of the ultimate partition do satisfy this condition

[3] The term "artificial intelligence" was coined by John McCarthy at a workshop held in Dartmouth in 1956.

[4] See Paul (1993) for a survey of approaches to abduction in the artificial intelligence literature. She also considers the knowledge-level approach as a further category, whereas I take this to belong to the logic-based approaches.

where Γ is identified with the reasoner's belief corpus. A further difficulty for artificial intelligence is that the term "abduction" is used for both a set of formulae satisfying the definition above and also the one chosen as the "best explanation" of the observations. This, as we have seen, is at odds with Peirce and Levi, who identify the process of selection with induction.

3.1. Restrictions on Abductions

At this point it is worth looking at certain restrictions placed on the selection of abductions. While these restrictions cannot determine precisely an ultimate partition, they can help to shape the character of the elements of the partition. The following restrictions are commonly considered.

Syntax. One of the most common restrictions is to impose some sort of syntactic restriction on the formulae constituting the background theory, observations, and abduction. In artificial intelligence the sentences in the background theory are commonly restricted to being Horn clauses. The main motivation for this is to make computation of abductions easier. As far as abductions are concerned, the main restriction is that they be conjunctions of literals. For Harman (1965), "[f]urthermore, the relevant explanations are always of the form *R because P, . . . , and Q*, explaining *why* or how it is that something is so. Achinstein (1983) points out that there are other sorts of abduction." In Reiter and de Kleer's (1987) Clause Management System, "[e]xplanations are conjunctions of ground literals" (p. 184).

Abducibles. Another very common restriction is to stipulate that abductions consist of propositional or predicate symbols from a defined set of such symbols (a subset of the language L). These symbols are termed *abducibles*. This is commonly used in logic programming (Kakas, Kowalski, and Toni 1995). A difficulty with this idea is to determine which symbols should be considered abducibles, especially since, as Stickel (1991) suggests, a suitable abducible for one inquiry may not be suitable for another.

Minimality and simplicity. Notions of minimality are often employed as a way of avoiding superfluous hypotheses. They are often motivated by appealing to Occam's razor. It is often applied in conjunction with other restrictions. For instance, adopting the syntactic restriction of limiting abductions to conjunctions of literals, a minimal abduction would be one where there is no other abduction containing a subset of its literals.

Triviality. In the sense defined above, an observation is an abduction for itself. Often, such an abduction is hardly compelling. To avoid this, we can specify that none of the observations should appear in the abduction. Alternatively, we can stipulate that the abduction make use of the background theory to avoid that the abduction should prove the observation alone (i.e., $\Psi \not\vdash \Phi$).

Specificity. When one abduction, together with the background theory, is able to prove another abduction, we say that the former is a *more specific abduction* than the latter. Stickel (1991) considers three types of abduction based on this idea. Most specific abductions are ones that contain literals that cannot be resolved with clauses in the background theory. Stickel claims that these are suited to diagnostic tasks. By assuming what we are trying to explain (i.e., the trivial abduction), we have the least specific abduction. These, according to Stickel, are suited to interpreting natural language.

3.2. *Selecting Abductions*

When it comes to selecting which abduction is "best," a large variety of usually quantitative methods is adopted. Since the literature here is quite extensive, we list only a few illustrative examples. Some of the methods used to select the "best" answer (abduction) include coherence (Thagard 1989; Ng and Mooney 1990), probability, cost (Charniak and McDermott 1985; Stickel 1991), and utility (Ram and Leake 1991). Interestingly enough, none of this work considers measures corresponding to Levi's insight that in abduction the reasoner is concerned with obtaining new information that is free from error.

3.3. *Abduction and Belief Revision*

In the area of belief revision, apart from the extensive work of Levi, there has been only a little work in the use of abduction. In Pagnucco (1996), I introduce a belief change operator known as *abductive expansion*, in which, instead of expanding solely with respect to the new information that a reasoner obtains, they first seek an explanation using abduction and expand by that instead (see the next section). However, I use the term "abduction" to refer to both the process of determining potential explanations and the choice of the "best."

Boutilier and Becher (1995) use the belief revision process itself to determine abductions. For them α *explains* β if and only if believing α causes the reasoner to believe β.[5] This is reminiscent of the Ramsey Test (Ramsey 1950) and suggestions by Gärdenfors (1990). In Levi's work, on the other hand, abduction is fundamental to the belief revision process.

[5] If β is believed at the outset, the reasoner can suspend judgement on β before determining whether it is explained.

In the artificial intelligence literature there is considerable debate as to the difference, if any, between abduction and induction. This appears to stem from a disregard for Peirce's view of abduction to which, as we have seen, Levi adheres. One problem is that induction (or enumerative induction) as described in the schema in section 1 can also be defined using the definition for abduction given at the start of this section. This is true, for instance, of the inverse resolution method proposed by Muggleton (1987). To differentiate the two processes one needs to resort to other criteria, as is often the case. One common distinction is that abductions, as determined in artificial intelligence, are often conjunctions of literals (cf. section 3.1), while enumerative inductions are often disjunctions of literals (i.e., clauses). Other distinctions are of course possible. Under Peirce and Levi's conception this problem is avoided and the distinction is clear. Insofar as both methods propose potential explanations, they are abductive. The point in the process where a choice of "best" explanation must be made is inductive.

4. DISCUSSION

> C. S. Peirce intially claimed that there were two kinds of ampliative inference: inductive and abductive; but he correctly recognized in later years that the types of ampliative inference he was calling "abductive" should be classified among the inductive inferences. (Levi 1996, p. 309)

In the artificial intelligence literature, debate has raged as to the difference between abduction and induction. It is fair to say that opinion is still very divided. In this section, I briefly consider an idea alluded to in Pagnucco (1996) although not well developed there (or elsewhere) concerning the nature of deduction, induction, and abduction. There I am dealing with the belief change framework introduced by Alchourrón, Gärdenfors, and Makinson (1985) (see also Gärdenfors 1988), which has become very popular in artificial intelligence.[6] As such, the formal setting differs from that of Levi, although the two still have many points in common.

In the AGM framework three forms of belief change are considered: expansion, contraction, and revision. Belief expansion is the process of adding new information to the reasoner's belief corpus without regard to error. Belief revision is the process of adding new information to the reasoner's belief corpus, removing beliefs if necessary to maintain consistency. Belief

[6] I henceforth refer to this as the AGM framework as is common usage in the literature.

contraction involves removing information from the reasoner's belief corpus (possibly for the purposes of suspending judgment or considering alternative beliefs).

In Pagnucco (1996) I consider an alternative to AGM expansion in which instead of adding the new information as is, an abduction for the new information is first determined and then added to the reasoner's belief corpus. This is closer in spirit to Levi's deliberate expansion, although the term "abduction" is used to refer to both the identification of potential answers and the selection of the "best" from among them. As a way of constructing such a function, I extend the epistemic entrenchment ordering of Gärdenfors and Makinson (1988). Epistemic entrenchment is essentially a ranked ordering (total preorder) over the reasoner's beliefs in which logical truths are maximally entrenched. Intuitively, epistemic entrenchment can be viewed as a preference over beliefs that determines what is retained and what is rejected during belief contraction. In my "abductive expansion" this ordering is extended over the nonbeliefs since these represent possible hypotheses (potential answers) and a method is given for selecting among them.

One final step that has been deemed missing in the AGM account is how the epistemic entrenchment structure evolves after the application of a belief change operation. This is known as the problem of *iterated belief change* (Nayak 1994; Darwiche and Pearl 1997; and Nayak, Pagnucco, and Peppas 2003). Several accounts have been put forward as to how best modify the epistemic entrenchment structure so that it is available for further contraction and (abductive) expansion as the problem exists in both cases.

As with Levi, belief change now boils down to a sequence of (abductive) expansions and contractions. The roles of the various inferential processes now become clearer. Abduction is concerned with the identification of potential answers, and induction selects the best of these, as suggested by Peirce and Levi. In this framework this is achieved by epistemic entrenchment. It might be suggested that epistemic entrenchment represents an ultimate partition although not strictly in Levi's usage of the term. Deduction determines the consequences of this choice. Enumerative induction dictates how iterated belief change is to occur. To be more precise, the mechanism chosen for determining how the epistemic entrenchment relation is iterated or evolves determines when and how the reasoner will generalize according to enumerative induction. On seeing one white swan, that all swans are white becomes a more plausible hypothesis but not sufficiently plausible to be deemed a potential answer (or at least not chosen by induction). A similar situation holds after seeing another white swan. After seeing numerous white swans, that all swans are white is now sufficiently high in the entrenchment ordering that it may be considered a potential answer and chosen by induction to be added to the reasoner's belief corpus.

5. CONCLUSIONS

At the abductive phase of inquiry . . . [a]ll that matters is whether the hypothesis proposed via abduction is sufficiently informative to merit serious attention.

(Levi 1986, p. 38)

This chapter has sought to look at Isaac Levi's ideas and thoughts on the conception of abductive reasoning. In particular, I have attempted to contrast Levi's work with that on abduction in artificial intelligence literature. It is a quite unfortunate state of affairs that Levi's deep insights into abduction have not been more widely considered in artificial intelligence. And it remains all the poorer for not having done so.

ACKNOWLEDGMENTS

I would like to thank members of the Knowledge Representation and Reasoning Program at the National ICT Australia and the University of New South Wales (and formerly the Knowledge Systems Group at UNSW and the University of Sydney) for the many discussions on various aspects in this chapter.

In particular, I would like to thank Isaac Levi, whose ideas have influenced many of my own thoughts on the subject of belief change and inference toward those ends. Each time I reread his works I discover many pearls of wisdom that had previously eluded me. I would also like to express a deep gratitude for much of the correspondence we have shared and Isaac's immense generosity in comments and ideas.

REFERENCES

Alchourrón, Carlos E., Peter Gärdenfors, and David Makinson. 1985. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions." *Journal of Symbolic Logic* 50: 510–30.

Allemang, Dean, Michael Tanner, Thomas Bylander, and John Josephson. 1987. "Computational Complexity of Hypothesis Assembly." In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pp. 1112–17. Los Altos, Calif.: Morgan Kaufmann.

Boutilier, Craig, and Veronica Becher. 1995. "Abduction as Belief Revision." *Artificial Intelligence* 77, no. 1: 43–94.

Charniak, Eugene, and Drew McDermott. 1985. *Introduction to Artificial Intelligence*. Reading, Mass.: Addison-Wesley.

Darwiche, Adnan, and Judea Pearl. 1997. "On the Logic of Iterated Belief Revision." *Artificial Intelligence* 89: 1–29.

Fann, K. T. 1970. *Peirce's Theory of Abduction*. The Hague: Martinus Nijhoff.

Gärdenfors, Peter. 1988. *Knowledge in Flux*: *Modeling the Dynamics of Epistemic States*. Cambridge, Mass.: MIT Press and Bradford Books.

Gärdenfors, Peter. 1990. "The Dynamics of Belief Systems: Foundations vs. Coherence Theories." *Revue Internationale de Philosophie* 44: 24–46.

Gärdenfors, Peter, and David Makinson. 1988. "Revisions of Knowledge Systems Using Epistemic Entrenchment." In *Proceedings of the Second Conference on Theoretical Aspect of Reasoning about Knowledge*, pp. 83–96. Los Altos, Calif.: Morgan Kaufmann.

Genesereth, Michael R., and Nils J. Nilsson. 1987. *Logical Foundations of Artificial Intelligence*. Los Altos, Calif.: Morgan Kaufmann.

Harman, Gilbert, H. 1965. "Inference to the Best Explanation." *Philosophical Review* 74: 88–95.

Kakas, Antonis C., Robert A. Kowalski, and Francesca Toni. 1995. "Abductive Logic Programming." In Dov M. Gabbay (ed.), *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 5. Oxford: Oxford University Press.

Levesque, Hector J. 1989. "A Knowledge-Level Account of Abduction." In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pp. 1061–7. Los Altos, Calif.: Morgan Kaufmann.

Levi, Isaac. 1979. "Abduction and Demands for Information." In I. Niiniluoto and R. Tuomela (eds.), *The Logic and Epistemology of Scientific Change*, pp. 405–29. North Holland: Acta Fennica Philosophica. Reprinted, with modifications, in Levi 1984.

Levi, Isaac. 1980. *The Enterprise of Knowledge*. Cambridge, Mass.: MIT Press.

Levi, Isaac. 1984. *Decision and Revisions*: *Philosophical Essays on Knowledge and Value*. Cambridge: Cambridge University Press.

Levi, Isaac. 1986. *Hard Choices*. Cambridge: Cambridge University Press.

Levi, Isaac. 1991. *The Fixation of Belief and Its Undoing: Changing Beliefs through Inquiry*. Cambridge: Cambridge University Press.

Levi, Isaac. 1996. *For the Sake of the Argument*: *Ramsey Test Conditionals, Inductive Inference and Nonmonotonic Reasoning*. Cambridge: Cambridge University Press.

Muggleton, Stephen. 1987. "Duce, an Oracle Based Approach to Constructive Induction." In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pp. 287–92. Los Altos, Calif.: Morgan Kaufmann.

Nayak, Abhaya C. 1994. "Iterated Belief Change Based on Epistemic Entrenchment." *Erkenntnis* 41: 353–90.

Nayak, Abhaya C., Maurice Pagnucco, and Pavlos Peppas. 2003. "Dynamic Belief Revision Operators." *Artificial Intelligence* 146, no. 2: 193–228.

Ng, Hwee Tou, and Raymond J. Mooney. 1990. "On the Role of Coherence in Abductive Explanation." In *Proceedings of the Eighth National Conference in Artificial Intelligence*, pp. 337–42. Los Altos, Calif.: Morgan Kaufmann.

Pagnucco, Maurice. 1996. "The Role of Abductive Reasoning within the Process of Belief Revision." Ph.D. thesis, Basser Department of Computer Science, University of Sydney.

Paul, Gabrielle. 1993. "Approaches to Abductive Reasoning: An Overview." *Artificial Intelligence Review* 7: 109–52.

Peirce, Charles Sanders. 1931–58. *Collected Papers*, vols. 1–8, ed. Charles Hartshorne, Paul Weiss, and Arthur Burks. Cambridge, Mass.: Harvard University Press.

Peng, Yun, and James A. Reggia. 1990. *Abductive Inference Models for Diagnostic Problem-Solving*. Berlin: Springer-Verlag.

Pople, Harry E. 1973. "On the Mechanisation of Abductive Logic." In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, pp. 147–52. Los Altos, Calif.: Morgan Kaufmann.

Ram, Ashwin, and David Leake. 1991. "Evaluation of Explanatory Hypotheses." In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, pp. 867–71. Hillsdale, N.J.: Erlbaum.

Ramsey, Frank P. 1950. *Foundations of Mathematics and Other Logical Essays*, 2nd ed. London: Routledge and Kegan Paul.

Reiter, Raymond, and Johan de Kleer. 1987. "Foundations of Assumption-Based Truth Maintenance Systems: Preliminary Report." In *Proceedings of the National Conference in Artificial Intelligence*, pp. 183–8. Los Altos, Calif.: Morgan Kaufmann.

Robinson, J. A. 1965. "A Machine-Oriented Logic Based on the Resolution Principle." *Journal of the Association for Computing Machinery*, 12: 23–41.

Stickel, Mark. E. 1991. "A Prolog-like Inference System for Computing Minimal-Cost Abductive Explanations in Natural-Language Interpretation." *Annals of Mathematics and Artificial Intelligence* 4: 89–106.

Thagard, Paul R. 1989. "Explanatory Coherence." *Behavioral and Brain Sciences* 12: 435–502.

11

Potential Answers – To What Question?

Erik J. Olsson

Throughout his career, Isaac Levi has advocated a unity of reason thesis according to which practical deliberation and theoretical inquiry, while differing as regards values and goals, are nonetheless similar from a structural point of view. In support of this thesis he has argued, first and foremost, that problems of induction can be seen as decision problems analogous to practical ones; both kinds of reasoning can be represented within the framework of Bayesian decision theory. In the practical case, the decision maker's options are practical actions. In the theoretical setting, they are, Levi thinks, acts of accepting potential answers to the inquirer's question. In this chapter I argue that his latter view exposes him to criticism unnecessarily.

1. LEVI'S COGNITIVE DECISION THEORY

Suppose that the agent is in an initial state of full belief *K*. There is a question that the agent wants to answer, and so he has identified a set of relevant propositions exclusive and exhaustive relative to *K* in the manner proposed by Levi in his first book from 1967 and defended in later works. This set is his ultimate partition *U*. Given *U*, a set of potential expansions relevant to the inquirer's demands for information can be identified. These potential expansions or, as Levi also calls them, *cognitive options* are formed by adding to *K* an element of *U* or a disjunction of such elements. The result is then closed under logical consequence. Levi assumes that adding some belief contravening proposition is also a potential expansion.

Suppose, for example, that we would like to determine the color of a given liquid after chemical reaction with another substance. If the experiment is difficult or costly to carry out, the agent might be interested in trying to settle the matter by predicting the outcome beforehand. Let us assume that from the agent's epistemic point of view, only three colors are seriously possible: red, white, or blue. The ultimate partition consists of "The color is red," "The color is white," and "The color is blue." The potential expansions are obtained by expanding K by one of the elements of the ultimate partition or a disjunction of such elements, that is, by "The color is red," "The color is white," "The color is blue," "The color is red or white," "The color is red or blue," "The color is white or blue," "The color is red or white or blue." Adding the last proposition would be refusing to expand beyond K, as the agent is already committed to accepting it. Finally, expansion by a belief-contravening proposition, such as "The color is both red and white," is also considered to be a cognitive option, in which case the inquirer will fall into inconsistency.

Now how are we to decide which potential expansion to implement? Levi's answer is that we should maximize the expected epistemic utility. The expected epistemic utility of adding a proposition A to the belief corpus is representable as $E(A) = \alpha Q(A) + (1 - \alpha)Cont(A)$, or some positive affine transformation thereof, where $Q(A)$ is the subjective ("credal") probability of A and $Cont(A)$ the informational value of A. Thus, the expected epistemic utility is a weighted average of the utility functions representing two different desiderata: probability and informational value. Moreover, $Cont(A)$ can be identified with $1 - M(A)$ where M is the information-determining probability of A. Information-determining probability is similar to Carnap's logical probability. Divide $E(A)$ by α and subtract from the result $q = (1 - \alpha)/\alpha$. The result will be $E'(A) = Q(A) - qM(A)$.[1] This is a positive affine transformation of the weighted average, and so maximizing this index is equivalent to maximizing the weighted average. The parameter q can be interpreted as the inquirer's degree of boldness. If q equals 0, the inquirer is an inductive skeptic; he will always refrain from expanding beyond the current corpus. A reasonable requirement is that falling into contradiction by adding a belief-contravening proposition should never be preferred to refusing to expand. Levi shows that this requirement is satisfied so long as $q \leq 1$.

To continue the example, suppose that Q("The color is red") $= 0.9$ and Q("The color is white") $= 0.09$, so that Q("The color is red or white") $=$ 0.99. Assume, further, that the information-determining probability assigned to each element of U is $1/3$. Hence, M("The color is red") $= 1/3$ and M("The color is red or white") $= 2/3$. Setting $q = 0.3$, the expected epistemic utility

[1] Here is an alternative path to this equation. In deliberate expansion, the decision maker evaluates the epistemic utility of expanding K by adding A without importing error and with importing error. The former is $1 - qM(A)$ and the latter is $-qM(A)$. The expected epistemic utility of adding A is determined by multiplying the first term by the credal probability that A and the second term by the credal probability that $\neg A$. Then the sum of these products is taken, which equals $Q(A) - qM(A)$.

of expanding by "The color is red" equals Q("The color is red") $- qM$("The color is red") $= 0.9 - 0.3 \cdot 1/3 = 0.8$. Consider now the weaker proposition "The color is red or white." Its expected epistemic utility will be Q("The color is red or white") $- qM$("The color is red or white") $= 0.99 - 0.3 \cdot 2/3 = 0.79$. Hence, expanding by "The color is red" has a greater expected epistemic utility than expanding by "The color is red or white." It is easy to verify that expanding by the former in fact maximizes expected epistemic utility. In this case, then, it was possible to settle the matter about the definite color of the liquid from an armchair position, without ever entering the laboratory.

It is not always the case, however, that a definite prediction can be made. A less bold agent, that is, one with a lower q-value, may not find the risk of predicting the wrong color worth taking. For instance, setting $q = 0.2$, the expected epistemic utility of expanding by "The color is red" equals 0.83, and that of "The color is red or white" equals 0.85. It can be verified that the optimal strategy, the one that maximizes expected epistemic utility, is to expand by "The color is red or white." While the agent can rest assured that the color is either red or white, he or she cannot say anything more definite about the color, and so he or she might be tempted to conduct the experiment after all.

As I believe the color example shows, it makes good sense to use Bayesian-style decision theory in a purely theoretical context. In recognition of the philosophical implications, Levi proposes a brand of pragmatism according to which "[w]hat is 'pragmatic'... is the recognition of a common structure to practical deliberation and cognitive inquiry in spite of the diversity of aims and values that may be promoted in diverse deliberations and inquiries" (1991, p. 78).

2. COGNITIVE OPTIONS BASED ON POTENTIAL ANSWERS

In Levi's theory, the set of cognitive options (potential expansions) relative to a question is the set of all expansions by some element of the ultimate partition or a disjunction of such elements. Semi-formally,

(1) An expansion $K + A$ is a cognitive option relative to a query Q if and only if A is an element of the ultimate partition or a disjunction of such elements.

The disjunctive structure is a feature that earlier decision theoretic accounts of induction lacked. For example, in Hempel's theory, which was devised independently by Levi in one of his early papers, the set of cognitive options was equated with the set of expansions by elements of the ultimate partition plus total suspension of judgment (Hempel 1960; Levi 1962). The possibility of partial suspense was not acknowledged. Thus, in the old Hempel-Levi theory, accepting that the color is either red, white, or blue would have been a legitimate option, whereas accepting that it is either red or white would not.

This is not all Levi has to say about the nature of cognitive options; he also believes that they should consist in accepting potential answers to the inquirer's question:

(2) An expansion $K + A$ is a cognitive option relative to a query Q if and only if A is potential answer to Q.

There is plenty of textual support for (2) in Levi's works where the expressions "cognitive option," "potential expansion," and "potential answer" are frequently used interchangeably, the result being that no clear distinction is made between (1) and (2). For instance, Levi commits himself to both in one swoop when he writes, recently, that "[t]he . . . potential answers (or the potential expansion strategies) of interest to the inquirer in the context of a given inquiry are [*sic*] linguistically representable as expansions of K by adding some sentence h in L equivalent given K to a disjunction of some subset of d_i's in U_K" (2004, sec. 2.2).[2]

As a consequence of (1) and (2), we have:

(3) A proposition A is a potential answer to Q if and only if A is an element of the ultimate partition or a disjunction of such elements.

Thus, the elements of the ultimate partition are potential answers to the inquirer's question. So, too, are all disjunctions of such elements.

It is, to be sure, reasonable to consider the elements of the ultimate partition to be potential answers to the inquirer's question. But in what sense is a disjunction of two or more elements of the ultimate partition also a potential answer? Surely, it cannot generally be the case that it is. If we are interested in determining the color of the liquid, we would be quite satisfied with learning that the color is blue, and also with learning that it is white. We would not, however, be entirely satisfied with knowing merely that the color is either red or white. If so, what is the basis for calling that disjunction a potential answer to our question? And what about the uncommitted disjunction expressing merely what we already knew at the outset, namely, that the color is red,

[2] U_K is the ultimate partition relativized to the corpus K. For a random sample from his earlier work, see Levi 1984, p. 92, where he contends that determining the cognitive options and identifying the list of potential answers are equivalent procedures.

blue, or white? Surely, we would be hard-pressed to call this an answer to our question, and yet, in the presence of (1), this is precisely what Levi's endorsement of (2), and hence also of (3), commits him to.

Despite its counterintuitive consequence (given (1)), Levi has continued to advocate and defend (2). Presumably, it plays a significant role in his theory, or it would not be worth retaining. My conjecture is that it has a part to play in his conception of the Peirce-Dewey pragmatist tradition, a tradition to which Levi pledges allegiance. Peirce, according to Levi, "conceived of induction as the task of eliminating potential answers from among those identified via abduction" (1991, p. 77). Levi may be understanding this conception to require that cognitive options be based on potential answers in the sense of (2). I return to this issue in the final section.

Given that (2) is worth retaining, the prima facie implausibility of (3) raises a problem for Levi. The strategy in the following is to reconstruct three possible defenses of (3) from Levi's writings.

3. THE "MODEST" DEFENSE

Let Q be the question that triggered the investigator's inquiry. Q is his main problematic – the question that motivates his investigation. For example, Q might be the question of whether the color of the liquid will be blue, red, or white after the chemical reaction has taken place. Given Q, we can formulate another question, the *modest question relative to* Q, denoted $Mod(Q)$: What is the strongest conclusion to be legitimately drawn while trying to answer Q, given the current evidence? For example, in the process of answering the question about the color of the liquid, the inquirer may wonder what he can conclude, at most, at the present stage of inquiry. In so doing he is, in my terminology, asking the modest question relative to his main problematic.

The first reconstructed argument refers to $Mod(Q)$ in order to support the disjunctive structure of the set of potential answers to Q. Its first premise runs as follows:

(4) A proposition A is a potential answer to $Mod(Q)$ if and only if A is an element of the ultimate partition (relative to Q) or a disjunction of such elements.

To this the following premise is added:

(5) A proposition A is a potential answer to Q if and only if A is a potential answer to $Mod(Q)$.

It is now pointed out that (4) and (5) together entail (3). So much for the bare logical structure of the argument. There can be no question about its validity. But are its premises true?

Let us start with (4). First, what are we to mean by a potential answer? The most natural way to think about a potential answer is as a proposition that, once it is accepted as true, terminates inquiry into the matter. Once a potential answer has been accepted, the inquirer no longer has any interest in pursuing the matter any further, and so he will not continue to look for new evidence – even if the costs of inquiry are negligible. To accept a potential answer is to remove the corresponding question from the research agenda.

Given this understanding of "potential answer," it should be clear that the set of potential answers to *Mod*(*Q*) does indeed have a disjunctive structure. The strongest conclusion to be legitimately drawn at a given point in time relative to a question *Q* may be any disjunction of elements of the latter's ultimate partition. In the worst case, it may be impossible to go beyond the current evidence. In that case, the answer to *Mod*(*Q*) is the uncommitted disjunction representing total suspension of judgment.

Suppose, for example, that *Q* is the question "What is the color of the liquid?" so that *Mod*(*Q*) is the question "What is the strongest conclusion to be legitimately drawn about the color of the liquid given the current evidence?" Surely, "The color is red or white" or any other disjunction is in principle a legitimate answer to the modest question. Such a disjunction, once accepted, gives us exactly what we asked for when *Mod*(*Q*) was raised, namely, the strongest conclusion to be legitimately drawn about the color given the present evidence. Every disjunction of elements of *Q*'s ultimate partition is in principle capable of being a complete answer to *Mod*(*Q*).

The truth of (4) is, I believe, taken for granted in Levi (1967), where he assumes that the sentences "eligible for acceptance as strongest via induction," which are in effect the potential answers to *Mod*(*Q*), are logically equivalent to those sentences that are members of a finite set of sentences *M* (p. 33). He goes on to stipulate that the set *M* contain all disjunctions of the members of the ultimate partition (relative to the original question *Q*; ibid., p. 34). This, of course, is just as it should be, given that he is in fact focusing on the question that I have called *Mod*(*Q*).

What reasons, then, are there for thinking that (5) is true as well? Why should we think that every potential answer to *Mod*(*Q*) is also a potential answer to the inquirer's original question *Q*? I think we have no reason to do so. On the contrary, we have every reason in the world to think that the potential answers to *Mod*(*Q*) that are mere disjunctions of elements of *Q*'s ultimate partition are *not* potential answers to *Q*. Learning that the color is either red or white may answer the question of what we can conclude at most about the color at the present state of inquiry, but it does not answer the original question that demanded the determination of a definite color.

Nonetheless, Levi does commit himself to (5) when he writes, in his 1967 book, that "the set of relevant answers to a given question is determined by the set of sentences in *L* that are eligible for acceptance as strongest via induction from the given evidence" (p. 33). On my reading, this statement entails that the set of potential or, as they were called at the time, relevant answers to the inquirer's original question *Q* is identical with the set of sentences that are eligible for acceptance as strongest via induction from the given evidence. The latter set clearly coincides with the set of potential answers to what I have called *Mod*(*Q*). But while Levi clearly endorses (5) in his 1967 book, he does not argue for it there; nor, to my knowledge, does he do so elsewhere.

Although there is clear textual evidence in Levi (1967) for attributing what I have called the modest defense of (3), there is also evidence pointing in the opposite direction. In the second half of the book, he observes that "it seems plausible to suppose that a complete answer would be obtained when an investigator is in a position to accept as true some element of *U*" (p. 143). An inquirer who has not yet accepted a member of the ultimate partition, by means of contrast, "will (provided the costs of inquiry are negligible) continue to look for new evidence, until he can justify a strongest consistent relevant answer to his question" (p. 145). Levi is here saying, in effect, that the class of complete potential answers to the inquirer's main problematic consists of all and only the elements of the ultimate partition. The inquirer is satisfied if and only if one of the elements of the ultimate partition has been accepted as the answer to his question. From this perspective, (5) must be false.

4. THE "ABDUCTIVE" DEFENSE

The strategy of justifying the disjunctive closure of potential answers by reference to modest questions seems gradually to have lost its importance in Levi's thought. In his latest book from 2004, there is no sign of his earlier idea of accepting something "as strongest." The emerging view seems to be that (3) is a fundamental principle that need not be justified in terms of modest questions or anything else. In Levi (1980), it is declared that "the condition that the set of potential answers be generated [in the disjunctive manner explained above] by an ultimate partition can be construed as a principle of abductive logic" (p. 46). In Levi (1984) it is listed as one of two principles of abductive logic (p. 93). The disjunctive structure of the set of potential answers, a property that that set intuitively seems to lack, is here presented as a brute fact. The present strategy is not really a defense of (3) in the sense of an argument in its support. The claim is rather that there is no need for any such argument, a claim that is contradicted by presystematic judgment.

In Levi (2004), we are told that "[t]he principles of abduction require, for example, that one be in a position to regard suspension of judgment between rival potential answers to be a potential answer" (sec. 2.2). The following explanation is furnished:

> I contend that we should not recognize as potential answers to the election question [i.e., the question of who will win an election, *X*, *Y*, or *Z*] that candidate *X* will win, that candidate *Y* will win or that candidate *Z* will win without also allowing as potential answers suspension of judgment between any pair or even all three. Philosophers tend to think of suspense or skepticism as an all or nothing affair. Either one suspends judgment between all elements of the ultimate partition or adopts one of them. But the tensions between belief and doubt are more nuanced than the aficionados of the battle between skepticism and opinionation would lead you to think. If someone insists that definite conclusions are to be recommended over suspense, that person should be required to show why suspense is inferior with respect to the goals of the inquiry and, if it is inferior, why some partial skepticism reflecting doubt between elements of some subset of the ultimate partition is not better than opinionation. Ruling out all or merely some forms of suspense as options by stipulation does not meet this demand.
> (sec. 2.2, notation adapted)

While this is indeed a compelling defense of the disjunctive structure of *cognitive options*, that is, of (1), it does very little in terms of giving an independent defense of (3). Rather, Levi seems to be presupposing that (2) is true. To be sure, the disjunctive structure of cognitive options supports the disjunctive structure of potential answers *provided that such options are based on potential answers*. But this is a trivial contention. The argument does not succeed in explaining how, contrary to appearance, (3) could be true. The net effect of the argument is instead to cast doubt on the tenability of (2), once it has been identified as a tacit premise.

5. THE "PARTIAL" DEFENSE

Hitherto "answer" has been taken to mean "complete answer." Let us by an *incomplete* answer mean a proposition that, while falling short of answering the question completely, still goes some way toward answering it. By a *partial* answer I mean one that is either complete or incomplete, that is, one that takes us *at least* one step closer to a complete resolution. The final line of defense takes (3) to be true provided that "potential answer" is taken to mean "partial potential answer."

To be sure, most disjunctions of partial answers are themselves partial answers in this sense. For instance, "The color is red or white" is not only a disjunction of partial answers; it is also a partial answer itself. In concluding that the color is red or white, we are in effect excluding the possibility of the color being blue, and so we are closer to obtaining a definite answer than we used to be.

Nonetheless, it does not hold generally that disjunctions of partial answers are themselves partial answers. The one exception is total suspense. While "The color is red or white" and "The color is white or blue" are both partial answers, this does not hold for their disjunction "The color is red, white, or blue." Accepting the latter means remaining in status quo, and so it obviously does not take us any closer to accepting a definite answer than we were before.

There are traces of the "partial" defense in Levi (1967). There he writes that "a relevant [potential] answer is, among other things, a sentence in *L* whose truth value is not decidable via deduction from the total evidence" (p. 33). Clearly, not being decidable via deduction from the total evidence is tantamount to being a partial answer in our sense. The qualification "among other things" presumably refers to the case of complete suspension that, although it is regarded as a potential answer by Levi, *is* decidable via deduction from the total evidence. On my reconstruction, Levi is here hinting at the interpretation of potential answers as partial answers while downplaying the fact that complete suspense does not fit into this picture.

6. ON THE PLACE OF POTENTIAL ANSWERS IN THE PRAGMATIST'S CONCEPTION OF INDUCTION

Why would Levi want to subscribe to (2) in the first place given that (3) follows once (1) is assumed true as well? I have already hinted that the answer may be sought in Peirce's account of induction or in Levi's conception thereof. According to Peirce, as Levi understands him, induction is the task of eliminating potential answers from among those identified by abduction. Levi might have been led to regard it as essential to the Peircean view that the cognitive option be based on potential answers in the sense of (2). This is the only explanation I can come up with for Levi's continued endorsement of (2) and his assigning it the status of a fundamental abductive principle in his later works.

Yet the satisfaction of (2) is in fact *not* required by the Peircean view. To see this, we recall that in all cases, except total suspense, employing Levi's inductive method means, in effect, rejecting the elements of the ultimate partition that are incompatible with the accepted proposition given our background knowledge. Hence, as Levi himself has often pointed out, his theory of induction can equivalently be described as a method for how, if possible, to reject elements of the ultimate partition. It follows that in order to achieve compliance with Peirce it is sufficient to require, unproblematically, that the elements of the ultimate partition be potential answers. It need not be stipulated, in addition, that disjunctions of such elements are potential answers.

In summary, in holding that all cognitive options correspond to accepting some potential answer to the inquirer's main problematic Levi exposes himself to severe criticism from the standpoint of common sense. If I am right, he does so unnecessarily: Neither the Peircean conception of inquiry nor any other element of Levi's theory that I have been able to identify depends crucially on the thesis that cognitive options should be based, in all cases, on potential answers to the inquirer's main problematic. Giving up the troublesome contention – which is the policy I recommend in the end – would, in Levi's jargon, not incur any significant loss of informational value.

ACKNOWLEDGEMENTS

I am indebted to Isaac Levi for our e-mail communications. His admirably patient responses to my sometimes opaque questions were instrumental in the eventual identification of the problem I raise here. I wish to thank Wlodek Rabinowicz for valuable comments on an earlier draft. Thanks also to Bengt Hansson and the participants of his philosophy of science seminar in Lund for their input.

REFERENCES

Hempel, C. G. 1960. "Inductive Inconsistencies." *Synthese* 12: 439–69.

Levi, I. 1962. "On the Seriousness of Mistakes." *Philosophy of Science* 29: 47–65.

Levi, I. 1967. *Gambling with Truth: An Essay on Induction and the Aims of Science*. Cambridge, Mass.: MIT Press.

Levi, I. 1980. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. Cambridge, Mass.: MIT Press.

Levi, I. 1984. *Decisions and Revisions: Philosophical Essays on Knowledge and Value*. Cambridge: Cambridge University Press.

Levi, I. 1991. *The Fixation of Belief and Its Undoing: Changing Beliefs through Inquiry*. Cambridge: Cambridge University Press.

Levi, I. 2004. *Mild Contraction: Evaluating Loss of Information Due to Loss of Belief*. Oxford: Oxford University Press.

12

Levi and the Lottery

Erik J. Olsson

> It is as rational to accept the hypothesis that ticket *i* will not win as it is to accept any statistical hypothesis that I can think of.
>
> *Henry E. Kyburg*[1]

> It is as rational to suspend judgement regarding the outcome of a fair lottery as it is to suspend judgement in any case I can think of.
>
> *Isaac Levi*[2]

1. INTRODUCTION

Consider a lottery of 1,000,000 tickets where there is one and only one winner and where one ticket is as likely as any other to be the winner. In this sort of scenario it is extremely improbable that any given ticket will win. The probability that a given ticket will lose is as high as that of any statistical hypothesis one can think of. If we are ever allowed to accept a hypothesis as true, then surely we are allowed to accept, of any given ticket, that it will lose. But to accept, for each ticket, that it will lose is to commit oneself to there being no winning ticket. This contradicts our background knowledge that there is a winning ticket. It seems, then, that common sense gives us a license for holding contradictory beliefs.

The lottery paradox, first formulated by Henry Kyburg, is still a hotly disputed subject that is thought to have all sorts of radical consequences for human inquiry.[3] Kyburg saw it as an argument for cultivating a tolerance for inconsistency and against demanding logical closure of a rational agent's beliefs. Some have taken it as a reductio argument against quantitative

[1] Kyburg 1963, p. 463.
[2] Levi 1965, p. 70.
[3] Kyburg 1961, pp. 196–7.

probabilistic accounts of inductive inference and as a positive reason for invoking qualitative methods such as default logic instead. Others have arrived at the opposite conclusion that it is the qualitative notion of full belief that is to be held responsible for causing paradox, urging that quantitative degrees of belief are all we need.

The main purpose of this chapter is to assess Isaac Levi's approach to the lottery.[4] Before I do this, I briefly say why I am dissatisfied with Kyburg's account and also with the related position taken recently by Luc Bovens and James Hawthorne (1999).

2. KYBURG ON THE LOTTERY

Given a theory *K*, Kyburg suggests that one can construct a new, more comprehensive corpus by adding to *K* all sentences that are sufficiently probable. In the case of the lottery, the inductively expanded set will be inconsistent. Kyburg argues that this need not be a regrettable fact, provided that the agent is not allowed to add conjunctions of sentences in the set to the set itself. The resulting set is, to be sure, inconsistent in the sense that every sentence of the language can be derived from it. But it is not inconsistent in another sense: No single one of the one million statements is inconsistent with the initial corpus. According to Kyburg, all we require of a set of sentences representing beliefs is that it not be inconsistent in the latter sense.

To most philosophers, though, it is bad enough that our beliefs are inconsistent in the sense of allowing everything to be derived. If the inquirer's belief set is inconsistent in this sense, then, for any given sentence *A*, both *A* and not-*A* will be derivable from the set. Since one of them is bound to be false, the agent is in certain error. It is difficult to see how it could be rational to enter such a clearly defective state of belief deliberately.

3. BOVENS AND HAWTHORNE ON THE LOTTERY

Let me first put Bovens and Hawthorne's account in perspective by relating it to Kyburg's. In his illuminating criticism, Isaac Levi confronts Kyburg's use of high-probability rules with the following trilemma: (1) One respects

[4] That the lottery paradox is a central concern in Levi's theorizing about inductive acceptance is witnessed by the frequency with which Levi returns to it in his writings. Here are some examples. The first treatment can be found in Levi 1965. The present description of Levi's theory is based on Levi 1967, pp. 38–41, 92–5. For a brief criticism of Kyburg's treatment focusing on the notion of acceptance, see Levi 1984, pp. 223–4, and Levi 1997, pp. 226–7. For a recent longer discussion, see Levi 1996, sec. 8.3.

the concern to avoid error in induction and rejects such rules, or (2) one abandons avoidance of error as a concern in inquiry, or (3) one denies that inductive acceptance removes doubt (see Levi 1996, p. 248). None of these three alternatives should be very attractive from Kyburg's point of view. This should be obvious as regards the first two horns of the trilemma.

Although Bovens and Hawthorne do not relate their theory to Kyburg's, their theory is, in effect, what Kyburg would get if he were to choose the third horn of the trilemma. For what they do is to combine a high probability rule with a notion of belief that does not require a belief to be free of doubt. For a person, they say, "*belief* is merely a convenient way to categorize those propositions for which her degree of confidence is no less than some threshold value q that she considers significantly high" (1999, pp. 245–6). Bovens and Hawthorne make clear that the threshold value q need not coincide with the maximum degree of confidence, that is, certainty (ibid., p. 246). That is to say, a person may, on their account, believe that a given proposition is true and yet not be entirely confident that it is true; some doubt may still remain as to its veracity. They emphasize that, as they use the term, categorizing something as a belief does not mean adding it to the stock of settled assumptions: "If asked whether she believes proposition S, α [the agent] may even explicitly report that her degree of confidence in S is no less than q, and that since she takes q to be an adequate threshold value for belief, she does indeed believe S" (ibid., p. 246).

The trouble with Kyburg's original position, as we saw, is that it licenses the presence of inconsistency in the set of settled assumptions. Given their reinterpretation of the concept of belief, Bovens and Hawthorne can consider it "rationally coherent" (ibid., p. 251) to believe both that some ticket will win and of every ticket that it will lose without thereby committing themselves to an inconsistent stock of settled assumptions. The reason, of course, is that the different lottery predictions are never added to that stock in the first place. Again, being a belief, on the present proposal, just means being assigned a sufficiently high degree of confidence.

This advantage, however, comes with a price tag. The difficulty facing Bovens and Hawthorne is to explain why in the first place the agent should take the trouble of separating from other propositions those propositions in whose veracity he or she has a relatively high degree of confidence. What is the point of this separation business? After all, the agent is not actually doing anything with the resulting set of propositions. There would, to be sure, be a point to it if the purpose were to add the separated propositions to the stock of settled assumption. But this, again, is a practice in which Bovens and Hawthorne do not want to engage.

In not admitting inconsistencies into the set of settled assumptions, Bovens and Hawthorne do distance themselves successfully from one of the less attractive features of Kyburg's original approach, that is to say, the legitimization of inconsistency among settled assumptions or full beliefs. But this benefit is attained at the cost of making beliefs irrelevant to inquiry and deliberation. In the final analysis, their proposal does not, in my view, represent a clear improvement on Kyburg's account.

4. LEVI ON THE LOTTERY

Let us see if Levi's own theory fares any better with respect to the lottery. As Levi sees it, inquiry starts with a question. Given a question, it may be obvious what the potential answers to that question are. If I ask in July 2004 who will win the next election for the American presidency, it is clear that the potential answers are "George W. Bush will win" and "John Kerry will win." In other cases, it is less obvious what the potential answers are. In such cases, there is need for what Levi, following Peirce, calls abduction, that is, the step in inquiry at which the potential answers are determined. Once the abduction step is completed, the stage is set for the identification of one answer as optimal in response to the inquirer's question. This answer is inductively acceptable. Levi sees the identification of an optimal answer as a cognitive decision problem analogous to a practical decision problem, and he recommends using what is in its essence Bayesian decision theory. In a final step, the optimal answer is added to the inquirer's stock of settled assumptions. This simplified account of Levi's complex theory will do for present purposes.

In Levi's view, inductive acceptability is relative to a question. What consequences does this have for the lottery paradox? Part of the formulation of the problem was that it appeared that all statements of the form "Ticket *i* will lose" are inductively acceptable. From the present perspective, this amounts to saying that "Ticket *i* will lose" is acceptable *in response to the inquirer's question Q*. So, what is the question in response to which "Ticket *i* will lose" is an optimal answer?

Most people confronted with the lottery would presumably simply ask which ticket will win. The potential answers to this question, however, are "Ticket 1 will win," "Ticket 2 will win," and so on. Statements of the form "Ticket 1 will *lose*" are not potential answers to this question. Hence, trivially, none of those statements can be an acceptable potential answer.

However, the matter is more complicated than it appears on first sight. Levi may object that I have not represented his theory correctly here. According to him, “Ticket 1 will lose” *is* a potential answer to the question of which ticket will win. That ticket 1 will lose is equivalent to the disjunction “Ticket 2 will win or ticket 3 will win or . . . or ticket 1,000,000 will win.” All such disjunctions of “strongest potential answers” are, in Levi’s view, also potential answers to the inquirer’s question. I have argued against this proposal elsewhere and will not repeat the point here.[5] Nonetheless, I do grant that adding “Ticket 1 will lose” to the body of evidence is one of the inquirer’s *cognitive options* in response to the question at hand, as indeed is adding any other disjunction of the same type. Adding it is surely a sensible reaction to the inquirer’s problem, and it does take him or her one step closer to a final solution, that is, to accepting a genuine potential answer. Anyway, what comes out of Levi’s decision theory once we consider accepting any disjunction of potential answers is that the optimal choice is to suspend judgment, that is, to accept only that some ticket will win but we do not know which one.[6]

But perhaps the inquirer is actually asking a whole series of questions. Perhaps he is asking of each single ticket at a time whether that ticket will win or lose. He is asking of ticket 1 whether it will win or lose, of ticket 2 whether it will win or lose, and so on. In general, he is asking of ticket i whether it will win or lose. If so, the potential answers are “Ticket i will win” and “Ticket i will lose.” Let us, following Rabinowicz (1979), call a question of this kind a *Hamlet question* (“To win or not to win . . . ”). On this reconstruction attempt, the statement “Ticket i will lose” does come out as a potential answer. This is promising news, although this does not by itself guarantee that it will also be an optimal potential answer.

Yet given its initial high probability, it seems plausible that it could be optimal as well. This is conceded by Levi.[7] That ticket 1 will lose may be the optimal answer to the question of whether it will lose or not; that ticket 2 will lose may be the optimal answer to the question of it will lose or not, and so on. What the lottery paradox illustrates, on this rendering, is that each statement in a set may be inductively *acceptable* even though the set as a whole is inconsistent with our background knowledge, for each statement of the form “Ticket i will lose” can be inductively acceptable relative to the corresponding Hamlet question, even though we are supposed to know that there is a winning ticket.

[5] See chapter 11 in this volume.

[6] For the details, see Levi 1967, pp. 92–3.

[7] “Thus, in the lottery problem, it is possible to predict that ticket 1 will not win if the ultimate partition U_1, which consists of the sentences ‘ticket 1 will win’ and ‘ticket 1 will not win,’ is used” (Levi 1967, p. 92). By the ultimate partition is meant the set of (strongest) potential answers.

Does this mean that the inquirer is justified in actually accepting all the statements in the troublesome set by adding all of them to the stock of settled assumptions, thus making that stock inconsistent? No, this does not follow, and in this I think Levi is right.

To see the point, consider again the statements of the form "Ticket *i* will lose" relative to their respective Hamlet questions. Suppose that the inquirer finds all these statements inductively acceptable and decides actually to accept them. Depending on how he or she attempts to accomplish this we get two subcases.

As for the first subcase, suppose that the inquirer decides to proceed stepwise, starting with ticket 1 and updating his or her belief set with the new item "Ticket 1 will lose," thus excluding ticket 1 from winning. As far as he or she can judge, the winning ticket could still be among the 999,999 remaining tickets. Accordingly, our inquirer now proceeds to the second ticket, noticing that its chance of losing is 999,998 in 999,999. As this is still a very high probability, our inquirer decides to exclude that ticket from winning as well, adding to the evidence that ticket 2 will lose.

However, it is not possible to continue excluding tickets in this way until no ticket is left. There will be a point at which further exclusion of tickets is no longer possible. The reason, obviously, is that as the inquirer accepts that a given ticket will lose, his or her evidence changes and with it the probability that a given remaining ticket will lose. This probability decreases.

I confine myself to illustrating the claim just made for a simple high probability rule. Suppose that the inquirer accepts a hypothesis if and only if its probability is at least 0.99 and that he or she has excluded all tickets but the 99 last ones. At this point our inquirer must stop, as the probability that the first of these remaining tickets will lose is $98/99 \approx 0.98989 \cdots < 0.99$. He or she cannot conclude that this ticket, or any of the other remaining ones, will lose, and so he or she need not worry about causing inconsistency in the set of full beliefs as the result of making different predictions.

As for the second subcase, suppose that the inquirer does not proceed by adding statements of the type "Ticket *i* will lose" to the evidence one at a time. Rather, when it has been observed that it would be rational to accept, of any single ticket, that it will lose, he or she decides to add all statements of that form to the evidence in one swoop. This would indeed make the evidence – the stock of settled assumptions – inconsistent, as he or she would then believe of each ticket that it will lose and at the same time believe that some ticket will win.

The impression of paradox vanishes once it is recognized that the last move is illegitimate. The inquirer should not be allowed, at a given point in time, to add several claims to his or her full beliefs, if those claims are answers to different questions. Thus, he or she should not be allowed to add, at a given time, both that ticket 1 will lose and that ticket 2 will lose, if the first statement is an answer to the question of whether ticket 1 will win or not and the second an answer to the question whether ticket 2 will win or not. These answers belong to different questions and hence to different inquiries. Nor should he or she be allowed to add as a new belief a conjunction of statements that are answers to different questions without first accepting each answer individually. The conjunction is not an answer to the current question. Either way, the lottery paradox suggests that pooling or simultaneous acceptance of inductively acceptable answers belonging to different questions should be forbidden.

5. HOW RESTRICTIVE ARE LEVI'S ASSUMPTIONS?

The two assumptions that make Levi's model immune to the lottery paradox are: (1) that inductive acceptability is relative to a question and (2) that one should not be allowed to pool answers to different questions by adding these answers simultaneously to one's stock of settled assumptions.

How serious are these restrictions? The first restriction – that acceptability is to be seen as relative to a question – does not seem serious at all. Human inquiry is always driven by a question. There is always an issue in response to which things are accepted or rejected. Hence, to borrow C. S. Peirce's celebrated metaphor, no roadblock of inquiry is introduced by assuming acceptance to be question-relative. A critic must show that it is important to human inquiry to have a notion of acceptance that is not question-relative. But it is difficult to see why anyone would care about inductive acceptance in the absence of an issue.

The second restriction – that the inquirer should not be allowed to pool answers to different questions in the manner described above – is less obviously an innocent one. I believe, however, that a strong case can be made for it. I fail to see how an inquirer could derive any advantages from pooling answers to different question as compared with an inquirer who proceeds in an incremental fashion, answering each question as it arises. If so, in forbidding pooling we are not obstructing the path of inquiry.

6. ON KNOWING THAT ONE'S TICKET WILL LOSE

On Levi's account, an inquirer may be justified in adding to his or her stock of full beliefs that a given single ticket will lose. In particular, he or she may add that the ticket drawn by him- or herself will lose. I do not think this is problematic in itself. Yet problems arise, I submit, when we combine this with Levi's theory of knowledge.

Levi has taken the controversial position that knowledge is just true belief. As soon as a person believes something fully and what he or she believes is true, he or she can rightly be said to know. Against this it may be objected that knowledge requires in addition that the knower has some sort of justification for his or her belief. Unimpressed by this type of criticism, Levi urges that an inquirer who already believes fully that something is the case is in no need of justifying that belief to him- or herself. After all, he or she is not in doubt regarding its truth.

This does not mean that an inquirer is never required to justify things to him- or herself. On the contrary, a rational agent is obliged to justify to him- or herself why a given belief *change* should be carried out. For instance, before adding a proposition to his or her full beliefs as an answer to a question, he or she is obliged to justify to him- or herself why that answer is better than other competing ones. Once the proposition has been accepted, there ceases to be a need for justification. There is no need for the inquirer to justify to him- or herself beliefs already held.

Now, as C. S. Peirce insisted, from the believer's own perspective all beliefs are true. If I believe something, then trivially I take what I believe to be true. Hence, "[f]rom *X*'s point of view at *t*, there is no difference between what he fully believes at *t* and what he knows at *t*" (Levi 1980, p. 28). Knowledge in the sense of true belief reduces to mere belief if the perspective taken is that of the believer him- or herself. If I fully believe that it rains, then, as far as I can judge, I know that it rains. By the same token, if I fully believe that my lottery ticket will lose, then, from where I stand, I know that it will lose.

But while one may claim to be certain that one's ticket will lose, in the sense of excluding it as a serious possibility, it seems awkward to claim to know that one's ticket will lose. Our lack of knowledge in this regard is as clearly a part of common sense as any other claim I can think of.[8]

The obvious reaction is to assign blame to Levi's already suspect minimal theory of knowledge. The suggestion would then be that we would do better if we were to adopt the standard justified-true-belief (JTB) analysis of knowledge. This, however, is not the case. Again, Levi's theory of induction allows that a person can predict, in the sense of adopting as a full belief, that his or her ticket will lose if, roughly speaking, losing is much more probable

[8] Harman and Sherman (2004) observe that "it is not intuitively correct that one can know using statistical reasoning that one's ticket is not the winning ticket."

than winning. Presumably, most defenders of the JTB analysis would say that in these circumstances the person is justified in believing that the ticket will lose. So, an inquirer's true belief that the ticket will lose may well be a case of *justified* true belief.

There are two main ways to react to this problem: (1) by rejecting the legitimacy of adding to one's full beliefs in single cases that a given ticket will lose or (2) by devising an alternative theory of knowledge. I have already said, in connection with Bovens and Hawthorne, why I find the first alternative unattractive. As I see it, the second path is the one to take. Knowledge is neither true belief nor justified true belief, if justification is understood, as it usually is, in terms of the individual's personal reasons for holding the belief in question.

According to the *social* account of knowledge that I favor, it is not sufficient for knowledge that the inquirer has his or her own personal reasons for believing what he or she does believe. Those reasons must be valid for others as well. Knowledge requires that the individual has a socially acknowledged right or even duty to believe in the circumstances in question. What he or she believes is what anyone in his or her position would believe as well. Wittgenstein put the matter as follows:

> When we say that we *know* that such and such, we mean that any reasonable person in our position would also know it, that . . . anyone endowed with reason . . . would know it just the same. (1977, sec. 325)

Beliefs arrived at via direct observation are of this kind: If I believe something as the effect of seeing it in broad daylight, then anyone standing where I were standing would arrive at the same belief. The same holds for beliefs based on clear memory and on the testimony of recognized experts on noncontroversial issues, just to mention two other traditionally celebrated sources of knowledge that also come out as such on the social view.

Contrary to what Levi maintains, then, when I say that I know, I am not just expressing my own personal certitude. I am also committing myself to the existence of grounds that are socially recognizable as such. I am giving others a license to take on the same view as I have. I am assuring them that there is no need on their part to bother with the details of justification. That part, I am promising, has already been taken care of. If it turned out that my belief was based on, say, reading tea leaves, I would, in claiming to know, be open to charges of misleading my audience.

The notion that knowledge is essentially social is certainly not new. We have seen that Wittgenstein held this view. Traces of it can be found in an early paper by Gilbert Harman in connection with his well-known newspaper example and in a recent paper by Robert Shope.[9] Levi does mention a social conception as an alternative to his own: "Of course, the claim that *X* knows that *h* at *t* may mean even for a pragmatist that *X* not only truly believes that *h* at *t* but does so authoritatively in the sense that in some way or other he can justify to others the truth of what he fully believes" (1997, pp. 67–8). In referring to others, this conception is clearly social in nature. As far as I can judge, Levi does not present any reasons for thinking that the social conception is inferior to his own individualistic analysis.

In urging that one should make a distinction between what one fully believes and what one knows, I am not introducing what Levi calls a "double standard of serious possibility."[10] Levi is right in insisting that what is seriously possible for a person is that which is compatible with the person's full beliefs. This raises the question of what role knowledge could play in inquiry and deliberation. I do not claim to have a complete account of what this role might be. I do tend to think that knowledge has a function in collective inquiries and, more specifically, in the efficient exchange of information between different inquirers. My knowledge is that part of my belief system that I, by means of making knowledge claims, can efficiently share with others. There is obviously more to be said about this, but that will have to await another occasion.

The social conception of knowledge qualifies as "pedigree epistemology," a term Levi uses for theories that are concerned with the origins of beliefs (e.g., Levi, 1980, sec. 1.1.). On the social view, whether or not a given belief counts as knowledge depends on how it was arrived at. Only belief acquisitions mechanisms that are common to all inquirers give rise to knowledge. Those that are peculiar to a given individual inquirer do not. If, as I have suggested, knowledge is useful because of its role in the efficient sharing of information among different inquirers, it makes sense to separate one's knowledge from one's mere beliefs. This process does require some concern with the pedigree of one's beliefs.

Levi thinks that pedigree epistemology is bad, urging that "[w]e ought to look forward rather than backward and avoid fixation on origins" (1980, p. 1). The pedigree theories he discusses in this connection are traditional foundationalism and what I have called JTB. Against the first, he objects that

9 Harman 1968; Shope 2002. Both Harman and Shope advance accounts of the social aspect of knowledge that differ from Wittgenstein's. I have defended a social conception in Olsson 2004.

10 In Levi's view, radical skeptics are guilty of employing double standards of serious possibility: an exclusive standard for the purposes of everyday life and an excessively liberal standard for philosophical purposes. Cf. Levi 1991, pp. 59–60.

"[t]here are no immaculate preconceptions" (ibid.). Against the second, he holds that there is no need to justify the truth of one's full beliefs to oneself ex ante (ibid., p. 28). I fail to see how any of this would count against the social theory of knowledge that I favor. Such a theory need not and, in my view, should not be associated with the notion of an incorrigible foundation. Nor is it part of such a theory that an inquirer needs to justify the truth of his beliefs to him- or herself after he or she has acquired them. Contrary to what Levi maintains, a concern with pedigree is not bad in itself – far from it – although he is certainly right in pointing out that some variations on that theme have little to recommend them.

What are the consequences of the social conception of knowledge for the lottery? Consider my full belief, arrived at by means of induction in response to the Hamlet question, that this ticket that I am holding in my hand (we pretend) will lose. Does this belief of mine satisfy Wittgenstein's social criterion? No. Surely, not everyone in my position – holding a ticket in his or her hand and contemplating whether it will win – would fully believe that the ticket will lose. If Wittgenstein's criterion were satisfied for this proposition, then no one would buy lottery tickets and there would be no lotteries. Lotteries exist because some people do not rule out the prospects of winning, although the likelihood of this happening is close to zero. These people, by the way, are not irrational; they are simply more cautious that those who rule out themselves being the winner. That inquirers are allowed to differ with respect to how cautious they are in their acceptances has been part of Levi's theory of induction from the start.[11]

Levi is committed to the counterintuitive view that a person can *know* that a given lottery ticket will lose. If I am right, the problem is to be located not in his theory of induction but in his account of knowledge that, although it was developed in opposition to mainstream epistemology, shares with that tradition of thought one of its main deficiencies, namely, an individualistic conception of knowledge. Once it is acknowledged that knowledge, as opposed to belief, is essentially social, the conclusion that a person can know that his or her lottery ticket will lose is not forthcoming. This, however, does not mean that a person cannot become certain that the ticket will lose. He or she can if the chances of winning are dim as compared with the chances of losing and the informational value of accepting that the ticket will lose is not terribly low. But this will be his or her own personal certainty based on grounds that are not recognized as such by the members of the community.

[11] The proposal to regard the degree of caution as a contextual parameter was made for the first time in Levi 1962.

Bovens, Luc, and James Hawthorne. 1999. "The Preface, the Lottery, and the Logic of Belief." *Mind* 108: 241–64.

Harman, Gilbert. 1968. "Knowledge, Inference and Explanation." *American Philosophical Quarterly* 5: 164–73.

Harman, Gilbert, and Brett Sherman. 2004. "Knowledge, Assumptions, Lotteries." *Philosophical Issues* 14: 492–500.

Kyburg, Henry, E. 1961. *Probability and the Logic of Rational Belief.* Middleton, Conn.: Wesleyan University Press.

Kyburg, Henry, E. 1963. "A Further Note on Rationality and Consistency." *Journal of Philosophy* 60 : 463–5.

Levi, Isaac. 1962. "On the Seriousness of Mistakes." *Philosophy of Science* 29: 47–65.

Levi Isaac. 1965. "Deductive Cogency in Inductive Inference." *Journal of Philosophy* 62: 68–77.

Levi, Isaac. 1967. *Gambling with Truth: An Essay on Induction and the Aims of Science*. Cambridge, Mass.: MIT Press.

Levi, Isaac. 1980. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance.* Cambridge, Mass.: MIT Press.

Levi, Isaac. 1984. *Decisions and Revisions: Philosophical Essays on Knowledge and Value.* Cambridge: Cambridge University Press.

Levi, Isaac. 1991. *The Fixation of Belief and Its Undoing: Changing Beliefs through Inquiry.* Cambridge: Cambridge University Press.

Levi, Isaac. 1996. *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inferences, and Nonmonotonic Reasoning*. Cambridge: Cambridge University Press.

Levi, Isaac. 1997. *The Covenant of Reason: Rationality and the Commitments of Thought.* Cambridge: Cambridge University Press.

Olsson, Erik J. 2004. "F. P. Ramsey on Knowledge and Fallibilism." *Dialectica* 58: 549–58.

Rabinowicz, Wlodek. 1979. "Reasonable Beliefs." *Theory and Decision* 10: 61–81.

Shope, Robert, K. 2002. "Conditions and Analyses of Knowing." In Paul K. Moser (ed.), *The Oxford Handbook of Epistemology,* pp. 25–70. Oxford: Oxford University Press.

Wittgenstein, Ludwig. 1977. *On Certainty*, ed. G. E. M. Anscombe and G. H. von Wright, transl. Denis Paul and G. E. M. Anscombe. Oxford: Basil Blackwell, 1977.

13

The Value of Truth and the Value of Information

On Isaac Levi's Epistemology

Hans Rott

1. A SERMON ON EPISTEMOLOGY

"I preach a sermon on epistemology," says Levi (2004, p. 77). Isaac Levi is an immensely interesting and powerful philosopher. In my opinion, he has erected a most impressive epistemological edifice in the past forty years. Inspection of current collections on epistemology, however, reveals that there is little interaction between Levi's work and almost any version of mainstream justification-based or reliability-based epistemology.[1] Why is this?

First, Levi resolutely distances himself from many main tenets of mainstream epistemology. He is very decidedly against any sort of "pedigree epistemology" (2004, pp. 11, 232), which holds that whether some belief counts as knowledge is dependent on its origin or its foundation, on where it comes from. He rejects the idea that convergence to the truth is the ultimate aim of inquiry (1980, pp. 70–2). And he is averse to "Parmenidean epistemology" (2004, pp. 10–12), according to which only logical, mathematical, or conceptual necessities should be admitted as *full* beliefs, while everything else should get assigned a degree short of "the Permanent One" (2004, p. 10).

For Levi, taking his position just means being true to the pragmatist stance. At any given point of time, a believer doesn't have to justify his or her currently held beliefs, since there is nothing other than the current set of beliefs on which the evaluation of the believer's mental state could be based (in Levi's terminology, there is no other "standard of serious possibility"). But that does not mean that believers are exempted from any duty of justification.

[1] In the recent *Oxford Handbook of Epistemology*, for instance, there is only one reference to Levi (Kaplan 2002, n. 25).

Believers have to justify their *changes of beliefs*. Justification in the pragmatist's sense means justification in terms of a believer's goals and values. Within the pragmatist camp, the *differentia specifica* of Levi's specific position is that justification should be decision-theoretic. As we shall see, truth and information figure most prominently in his own decision-theoretic account.

Second, even the term "knowledge" has come to play a minor role in Levi's works. In *The Enterprise of Knowledge*, he has only a short discussion of the justified-true-belief analysis (1980, pp. 1–3, 28–30), and in *The Fixation of Belief and Its Undoing* (1991, p. 45), he rather casually says that knowledge, as he uses the term, is "error-free, full belief," and he dismisses questions of justification as "irrelevant." I have not been able to find any definition or analysis of knowledge in his most recent books, *For the Sake of Argument* (1996) and *Mild Contraction* (2004). To put it provocatively, it begins to seem as if Levi is preaching a sermon on epistemology without knowledge.

A third reason for the unfortunate neglect of Levi's work by epistemologists may lie in the fact that much of Levi's presentation is rather technical. It can be hard to follow his philosophy without attending to a lot of logical and probabilistic niceties. Levi finds large audiences among researchers in philosophical logic, artificial intelligence, and knowledge representation, but he has not been equally successful in getting his message across to epistemologists who are less used to dealing with technicalities.

Before venturing a critical evaluation of Levi's work, I shall recapitulate as perspicuously as I can some central elements, paying special attention to their most recently renovated forms. It is mainly the current state of Levi's epistemology as expounded in his new book *Mild Contraction* on which I focus.

2. INQUIRY

Levi's epistemology focuses on the notion of inquiry. *Inquiry*, according to Levi, consists in the answering of questions and the solution of problems (1996, p. 161; 2004, pp. 39, 56). It has as its proximate cognitive aim the change of bodies of belief or theories with a view to gaining new *error-free information* (1996, p. 165; 2004, pp. 70–1, 76–80). Levi's models can be seen as attempts to implement in a precise and formally explicit manner James's famous double maxim: "Seek valuable information! Shun error!"[2] As we shall

[2] Levi (1991, p. 81; 2004, pp. 76–80). In James's (1979, p. 24) own words: "*We must know the truth*; and *we must avoid error* – these are our first and great commandments as would-be knowers." Levi rightly points out that "truth" here should be replaced by "information" or "valuable information" and that James's double maxim was anticipated by Peirce in his *Harvard Lecture X* held in 1865.

see, one has to be clear about the fact that the two parts of James's maxim give expression to two very different desiderata.

So inquiry can be understood as the process of rational belief change. By *beliefs*, Levi refers not to conscious occurrences in the inquirer's mind but to his or her commitments. This reading of "belief" motivates Levi's imposing the constraint that the set of the inquirer's beliefs be logically closed.

Levi formalizes the process of inquiry. It is not possible to understand fully his intent without going through his formal setting for the analysis of the norms of rational or scientific deliberation. In this section I recapitulate the formalization in its barest backbones.

$\underline{K}$ denotes the class of all potential corpora of beliefs ($\underline{K}$ is sometimes called "conceptual framework"; 1996, p. 18; 2004, p. 41). For Levi, a potential belief state $\boldsymbol{K}$, abstractly conceived, can be represented by a potential corpus K in $\underline{K}$. While a belief state is a nonlinguistic entity, a corpus is a set of sentences in a regimented language L. Since K is supposed to represent the "full beliefs" of an inquirer, that is, his or her "commitments" or "standards of serious possibility," K is supposed to be deductively closed (1996, p. 13; 2004, pp. 14, 41–2). Usually, but not always, K is assumed to be consistent.

To avoid undue technicalities, I shall in this chapter suppose that the set of all L-sentences is partitioned by the relation of logical equivalence into finitely many equivalence classes. As a consequence, each potential corpus K can alternatively be thought of as being represented by a single sentence, viz., the conjunction of representatives of all equivalence classes included in K. I shall take the liberty of jumping between both types of representations of belief states as convenience suggests.

The *ultimate partition* U is a partitioning of the space of all possibilities, where possibilities are identified with maximally consistent sets of L-sentences. Alternatively, U can be thought of as a set of pairwise incompatible and jointly exhaustive L-sentences. Given a corpus K, we mean by the elements of the partition *within* K the classes in U whose maximally consistent sets of L-sentences include K (the first way of thinking of possibilities) or, alternatively, the sentences in U that imply K (the alternative way of thinking). The elements *outside* K are just those elements of U that are not within K.

For the purposes of this chapter, we assume that the content of any potential belief state is expressible as the intersection of the elements of certain sets in U or, alternatively, as the disjunction of certain sentences in U. We also assume that all potential answers to questions in which the agent may be interested, all his or her demands for information are expressible as combinations – intersections or, alternatively, disjunctions – of elements of U (1980, p. 45; 1996, p. 163; 2004, p. 49). The ultimate partition determines, among other things, the granularity of currently and potentially held beliefs. Inquirers may alter their ultimate partitions for some reason or other, but such cognitive operations are not a main subject of Levi's studies.[3]

Levi imposes a *commensuration requirement* (1991, p. 65; 2004, p. 16), according to which any legitimate belief change is decomposable into a series of belief expansions and belief contractions. So what inquirers need are principled methods for legitimately expanding and legitimately contracting their corpora.

An *expansion* of the corpus K by a hypothesis h is denoted by $K + h$, a *contraction* of K with respect to a hypothesis h is denoted $K \dot{-} h$. According to Levi, there are no other legitimate ways of modifying corpora besides expansions and contractions. A *belief revision* by an hypothesis h, for instance, gets reduced to a compound consisting of a contraction followed by an expansion, $K * h = (K \dot{-} \sim h) + h$. This equation has become widely known as *the Levi identity*.

3. THE AIMS OF INQUIRY: TRUTH AND INFORMATION AS COGNITIVE VALUES

For Levi, rational belief change has to be justified in decision-theoretic terms. To see how this works, we have to supply some more formal means.

If K is the current corpus, then the *credal probability* relative to K is Q_K. All full beliefs, that is, all elements of K, get credal probability 1. The expansion $K + h$ gets credal probability $Q_K(K + h) = Q_K(h)$ relative to K (1996, p. 167). Credal probabilities are also called "*expectation-determining probabilities*," "*belief probabilities*," or "*confirmational commitment*" (1996, pp. 167–9; 2004, pp. 78, 84, 89).

The measure of informational value is similarly functionally dependent on what happens to be the inquirer's current corpus. The basic structure is encoded in an informational-value determining probability function M_K associated with the belief state K. The *informational value* of a sentence h is determined to be $1 - M_K(h)$, where M_K is the *informational-value determining probability function*: $Cont_K(h) = 1 - M_K(h) = M_K(\sim h)$.[4] All full beliefs

[3] We may neglect for our purposes Levi's *basic* or *minimal corpus LK* with the accompanying partition U_{LK} (2004, pp. 49, 57). Levi introduces LK to represent the minimal presuppositions of the inquiry and U_{LK} to represent the maximally specific relevant answers given these presuppositions (regardless of how K changes). The ultimate partition is similar in function to Shafer's (1976) *frame of discernment*. Cf. Fioretti 2001.

[4] Levi (1967, pp. 69 and 107; 1991, p. 84; 1996, pp. 24 and 169; 2004, p. 84). This function has been suggested as a measure of information or content at least since Carnap and Bar-Hillel

get informational-value determining probability 1, that is informational value 0.

Levi (2004, pp. 98–101) mentions other content measures, such as $1/M_K(h)$ and $-\log M_K(h)$. The content measure chosen by Levi has a number of advantages; $Cont_K$ values, for instance, have a finite upper and lower bound. More important, they satisfy the condition of *Constant Marginal Returns in Informational Value of Rejection*, which says that the informational value gained by rejecting an element x of the ultimate partition (thereby expanding the inquirer's corpus) is independent of the set of further elements rejected along with x.[5] But the most important advantage is, as we shall see, that this content measure $Cont_K$ can elegantly be weighed against the inquirer's credal probability Q_K.

In Levi's early writings, M_K is assumed to be the traditional "logical probability," so if there are four possibilities between which the agent is to choose, the M_K-value of the hypothesis that a particular one is true would be $^1/_4$ and its $Cont_K$-value would be $^3/_4$. Levi insists that informational-value determining probabilities are in general *not* identical with credal probabilities (1996, p. 22; 2004, p. 88).[6]

There remains an unresolved problem of interpretation: Where do information-determining probabilities come from? Clearly, informational-value determining probabilities do not represent subjective beliefs, nor frequencies, nor propensities. I suspect that the only way to understand what they mean is via their intended interpretation as information-value determining. But then it seems a legitimate question to ask whether it would not be clearer to use *informational value* as the primitive notion, without a recourse to uninterpreted probabilities.

We may conveniently summarize what is relevant to inquiry according to Levi – the set of "contextual parameters" (2004, p. 89) – in a quadruple of the form $\langle K, U, Q_K, M_K \rangle$. Levi never introduces such 4-tuples formally as devices

(1952, p. 237). *Cont*(h) is independent of the factual truth of h. Measures of informational value are discussed in detail in Levi (1969).

[5] Levi (2004, pp. 83, 100–1). It is evident from Levi's wording that he is aware that the principle does not transfer well to his theory of belief *contractions*: "The following assumption (proposed in 1967 in Levi, 1984, ch. 5) works very well *in the context of expansion*" (2004, p. 83, my emphasis). Also compare Levi's condition of *Extended Weak Positive Monotonicity* (2004, p. 126).

[6] Levi also discusses the idea that the two probability functions Q_K and M_K associated with the inquirer's current belief state K are obtained by conditionalization from some "master probability functions" Q and M associated with the basic belief state LK (2004, p. 153; cf. 1996, pp. 167–9, 267).

for the analysis. For ease of reference, however, let us call them *frameworks for inquiry*. Most of such a framework for inquiry seems subjective, relative to the inquirer in question, but, as I said, the status of informational-value determining probabilities is unclear (to me).

Let us now see how Levi applies frameworks for inquiry in processes of belief change.

4. AGGREGATING TWO VALUES IN DELIBERATE INDUCTIVE BELIEF EXPANSIONS

Levi's account of deliberate inductive expansion is a beautiful piece of theorizing, but it is not easy to understand. Let us develop it carefully.

One of the central problems addressed by Levi is the problem of how to extend a given corpus through inductive reasoning. Such an extension aims at improving one's doxastic state, without any "external disturbances" through new evidence. The question is which new hypotheses to accept. According to Levi, inquirers should not try to maximize or satisfice probabilities, nor should they look for some tailor-made inductive or nonmonotonic logic. They should rather use decision theory. In Levi's view, the traditional notion of a *justified expansion* reduces to (or is to be replaced by) the notion of *a utility-maximizing expansion*. Inductive reasoning is an exercise in decision theory.

Let the framework $\langle K, U, Q_K, M_K \rangle$ be given. Each hypothesis h suitable for strengthening K can be represented as a disjunction $x_1 \vee \cdots \vee x_n$ of elements of the ultimate partition *U within K*.

Levi's principal pragmatist idea is to assess hypotheses according to their cognitive values. The cognitive value of an hypothesis h is a weighted average of *the value of its truth* and *the value of its informational content*.

$$\text{Value}(h) = \alpha \cdot \text{T-value}(h) + (1 - \alpha) \cdot \text{I-value}(h)$$

The parameter α expresses the relative importance that is attached to the truth of a hypothesis, and $1 - \alpha$ expresses the relative importance of its informational value. α ranges between 0 and 1, and its particular value is chosen by the inquirer. If $\alpha = 0$, then truth does not matter at all, only the amount of information conveyed by h counts (so tautologies are regarded to be of minimal value). If $\alpha = 1$, then information does not matter at all, only the truth of h counts (so tautologies are regarded to be of maximal value). The more cautious the inquirer, the larger is α; the bolder the inquirer, the smaller is α. Proceeding from the "plausible assumption that no error is to be preferred to any case of avoiding error" (1967, p. 107; 1996, p. 171; 2004, p. 86), Levi holds that α should not be less than 0.5.

Now Levi makes a number of important decisions. Both T-value and I-value are "normalized" so that they take values between 0 and 1. The values of truth and falsity are supposed to be *identical* for all possible hypotheses: If h is true, then $\text{T-value}(h) = 1$, if h is false, then $\text{T-value}(h) = 0$.[7] As already mentioned, the value of information is measured by $Cont_K(h) = M_K(\sim h)$. We see below that these particular ways of defining the value of truth and the value of information have important consequences for the formal structure of Levi's theory.

The value of accepting h depends on the value of α (something subjectively chosen) and the truth value of h (something beyond the agent's control). Making explicit this double dependence, we can write down the value of accepting a hypothesis h as

$$\begin{aligned}
\mathrm{V}_\alpha(h, \text{true}) &= \alpha \cdot \text{T-value}(h, \text{true}) + (1-\alpha) \cdot \text{I-value}(h) \\
&= \alpha \cdot 1 + (1-\alpha) \cdot (1 - M_K(h)) = \alpha + (1-\alpha) \cdot (1 - M_K(h)) \\
\mathrm{V}_\alpha(h, \text{false}) &= \alpha \cdot \text{T-value}(h, \text{false}) + (1-\alpha) \cdot \text{I-value}(h) \\
&= \alpha \cdot 0 + (1-\alpha) \cdot (1 - M_K(h)) = (1-\alpha) \cdot (1 - M_K(h))
\end{aligned}$$

(This differs a little from Levi's own notation.) Now the inquirer often does not know whether h is true or false. In this case, he or she may employ his or her credal probabilities Q_K about the hypothesis's truth value. Decision theory says that the inquirer ought to maximize his or her expected utility:

$$\begin{aligned}
\mathrm{EV}_\alpha(h) &= Q_K(h) \cdot V_\alpha(h, \text{true}) + Q_K(\sim h) \cdot V_\alpha(h, \text{false}) \\
&= Q_K(h) \cdot (\alpha + (1-\alpha) \cdot (1 - M_K(h))) + (1 - Q_K(h)) \\
&\quad \cdot((1-\alpha) \cdot (1 - M_K(h))) \\
&= \alpha \cdot Q_K(h) + (1-\alpha) \cdot (1 - M_K(h)) \\
&= (1-\alpha) + \alpha \cdot Q_K(h) - (1-\alpha) M_K(h) \\
&= (1-\alpha) + \alpha \cdot \Sigma_{\{x \text{ in } U : x \vdash h\}} Q_K(x) - (1-\alpha) \cdot \Sigma_{\{x \text{ in } U : x \vdash h\}} M_K(x) \\
&= (1-\alpha) + \Sigma_{\{x \text{ in } U : x \vdash h\}} (\alpha \cdot Q_K(x) - (1-\alpha) M_K(x))
\end{aligned}$$

In the transition to the last line, we can see how important it is that Levi chooses the particular definitions of the values of truth and information.

[7] Levi (1967, p. 106; 2004, p. 80). This, of course, is a very strong idealization. It neglects all matters of relevance. Even if they are equally likely, the truth that the Dallas Mavericks won their last match matters much less to me than the truth that my daughter got home safely last night. Intuitively, I'd say that these truths differ a lot *in value* for me.

Different concepts of T-value or I-value would not have allowed him to break down the value of accepting h into a sum of values of accepting those elements of the ultimate partition that entail both K and h.

What the inquiring subject should do, in Levi's pragmatist picture, is to accept an hypothesis h that maximizes the value of $\mathrm{EV}_{\alpha}(h)$. But such a hypothesis is easy to find. We just have to collect all those x in U within K for which the last term in the formula above is nonnegative, that is, for which

$$Q_K(x)/M_K(x) \geq (1-\alpha)/\alpha$$

If this inequality is satisfied, then x is a disjunct of the hypothesis to be accepted. Decision theory is actually silent about whether one should take an x for which the value $Q_K(x)/M_K(x)$ is exactly zero. Levi recommends not to reject zero-valued elements. His *Rule for Ties in Expansions* (1967, p. 84; 1991, p. 93; 1996, p. 172; 2004, p. 87) instructs the inquirer to take the *weakest* optimal expansion if there is one. And there always is one, namely, the one that we obtain by disjunctively conjoining the zero-valued elements to the positively valued elements. We have now arrived at Levi's *Inductive Expansion Principle* (see 1967, p. 86; 1996, p. 172; 2004, pp. 87–8).

Define the *index of boldness* $q = (1-\alpha)/\alpha$. Note that the value of q increases as the inquirer's degree of caution α decreases. Since α is supposed to range from $^1/_2$ to 1, q ranges from 0 to 1. For every element x of the ultimate partition, the chances that x is rejected as a result (as a "conclusion") of an inductive inference increase as the index of boldness q is increased. If q is raised, this means that fewer elements of U within K will be left uneliminated, hence the selected hypothesis h will contain *more* information.[8]

Now let us change perspectives. Instead of taking a certain number α (or equivalently, a certain number q) as given and asking which hypothesis should be accepted, take a certain member of the ultimate partition as given, and vary α (or q). Given the inquirer's credal probability Q_K and information-determining probability M_K, there is for each x in the ultimate partition within K a unique number

$$q(x) = \min\{Q_K(x)/M_K(x),\ 1\}$$

that is just low enough to make x a disjunct of the hypothesis to be accepted. If the index of boldness q chosen by the inquirer is higher than $q(x)$, then the possibility that x is true is ruled out (that is, $\sim x$ is accepted). If q is lower than

[8] One could also think of inserting the index q of boldness as a fifth element into a framework of inquiry. But in contrast to the elements we have identified above, q seems to be a matter of the inquirer's free choice, so I prefer to leave it out.

or identical to $q(x)$, then x remains a serious possibility. $q(x)$ is "the maximum value of q at which x fails to be rejected" (Levi 1967, p. 137; 1996, p. 185; 2004, pp. 89–90).[9]

That an element x of the ultimate partition is *rejected* means that $\sim x$ is one of the agent's full beliefs (assuming that the elements of an ultimate partition are expressible by sentences of the agent's language). The more elements are rejected, the stronger are the inquirer's beliefs. So if q is higher than $q(x)$, then $\sim x$ gets accepted; if q is less than or equal to $q(x)$, then $\sim x$ does not get accepted. The degree of boldness q that the inquirer chooses has to be higher than $q(x)$ in order to make $\sim x$ acceptable. The higher $q(x)$, the more boldness or "mental effort" it takes to find $\sim x$ acceptable. If $q(x)$ is less than 1, it is possible, by a sufficient amount of boldness, to accept $\sim x$. But if $q(x)$ equals 1, this is never possible, not even by the utmost exertion of one's boldness. In a way, one could say that $q(x)$ is the degree of nonbelief of $\sim x$.

This idea can be generalized to propositions that are not elements of U within K. Hypotheses h that are not maximally specific are identified with disjunctions $x_1 \vee \cdots \vee x_n$ of those elements $x_1, \cdots, x_n$ of the ultimate partition that imply h. A hypothesis h equivalent with $x_1 \vee \cdots \vee x_n$ is *rejected* if all the x_i's are rejected (1996, p. 185).[10] Thus, clearly the maximum level of q at which h fails to be rejected, $q(h)$, equals the maximum of the q-values of the disjuncts x_i:

$$\text{if } h \text{ is equivalent to } x_1 \vee \cdots \vee x_n \text{ within } K\text{, then } q(h) = \max(q(x_i))$$

In my alternative way of speaking, $\sim h$ gets accepted, if the q-value chosen by the inquirer is higher than $q(h)$; otherwise, $\sim h$ will not be accepted. Again, it is possible to find $\sim h$ acceptable at some level of boldness if and only if $q(h)$ is less than 1. On the other hand, $q(h)$ is 0 if and only if $\sim h$ is already accepted in K.

The *Shackle degree of belief* of h (relative to a framework of inquiry $\langle K, U, Q_K, M_K \rangle$) is defined to be (Levi 1996, pp. 185–6; 2004, p. 90[11])

$$b(h) = 1 - q(\sim h)$$

9 If x is rejected for all (no) values of x, then Levi sets $q(x) = 0$ (resp., $q(x) = 1$).

10 I find this terminology slightly confusing. On the one hand, h is here thought to be a candidate belief (and not an actual belief), so in a sense, it does not have to be rejected in the first place. On the other hand, the *proposal to expand* the current corpus K by a hypothesis $h = x_1 \vee \cdots \vee x_n$ would probably be rejected as soon as *one* (and not *all*) of the disjuncts x_i turned out to be rejectable. So instead of saying that h gets rejected if all x_i's are rejected, I would find it clearer to say in this case that $\sim h$ is accepted.

11 Actually, Levi (2004) inserts an intermediate concept, viz., the Shackle *degree of potential surprise* or *degree of disbelief* $d(h) = 1 - q(h)$, and then defines $b(h) = d(\sim h)$.

The more boldness it requires to render h acceptable, the *lower* its degree of belief is. This is why the q-values are subtracted from 1. The b-value of a proposition h is 1 if and only if it is in the inquirer's corpus K. It is 0 if h cannot be found acceptable for any degree of boldness (this is true either for h or $\sim h$, for every hypothesis h).

Levi summarizes his achievement as follows: "[W]e have derived a measure of degree of disbelief and degree of belief exhibiting the formal properties of Shackle measures of degrees of potential surprise or disbelief and of degrees of belief. The derivation involves the use of a family of deductively cogent, caution-dependent, and partition-sensitive criteria for inductive expansion" (1996, p. 186).

It is this construction that most intimately ties Levi's decision-theoretic, pragmatist philosophy together with the work done on the logics of belief revision in the AGM paradigm and related models. An important role in Levi's account of both deliberate expansion and contraction[12] is played by Shackle-like measures, so-called after the British economist G. L. S. Shackle (1949). Shackle measures assign values to propositions, and Levi presents them in a normalized form with a range between 0 and 1. The characteristic property (1967, p. 133; 1996, pp. 181, 264; 2004, p. 90) of a Shackle measure b is that for all propositions g and h

$$b(g \wedge h) = \min\{b(g), b(h)\}$$

According to Levi (1991, p. 182; 1996, pp. 180, 258; 2004, pp. 90–5), variants of Shackle measures have been rediscovered or reinvented in the last three decades by many researchers, including L. Jonathan Cohen, Didier Dubois and Henri Prade, Peter Gärdenfors and David Makinson, and Wolfgang Spohn.

In my opinion, Levi's decision-reconstruction of belief functions (in Shackle's sense) is a remarkable achievement. Starting with his early masterpiece *Gambling with Truth*, he has developed an ingenious model of combining the demands for truth and information. He has shown how to take the decision-theoretic prescription of utility maximization as the primary principle, how to apply it to the important epistemological problem of inductive expansion, and how to reach an unequivocal decision by the Rule for Ties, which recommends adopting the logically weakest solution of the

[12] Levi insists that the *kind* of use Shackle measures are put to is very different in deliberate expansion from that in contraction (1996, pp. 267–8; 2002, pp. 136–7; 2004, pp. 40–1). I address contractions in section 6, but unfortunately, I am not able to deal with this question in the present chapter.

maximization problem. Moreover, Levi has made clear how the problem of maximizing epistemic utility (by a suitable choice of a hypothesis h) can at the same time be viewed as a satisficing problem with respect to a measure of belief: The agent can accept every possibility above a certain "level of aspiration" or threshold value q.

In a more critical vein, it remains to say that the particular definition of the I-values is not very well motivated. Several questions remain. Where does M_K come from? Why exactly is the content measured by $1 - M_K$ rather than, say, by $1/M_K$ or $-\log M_K$? A similar complaint may be raised against the definition of the T-value. What is the justification for assigning truth the constant value 1 and assigning falsity the constant value 0, irrespective of the relevance of the proposition in question?

5. THE SINGLE VALUE FOR BELIEF CONTRACTIONS: INFORMATIONAL VALUE

Like belief expansions, belief contractions are considered to pose decision-theoretic problems by Levi. Prima facie, the problems presented by contractions, that is, by retractions of belief, are completely dual to the problems presented by expansions. The positive aspect of an expansion is that one gains information; the negative side is that in gaining information one may import error.[13] We have seen how Levi conceives of the trade-offs between these two factors. The negative aspect of a contraction is that one loses information; the positive side, it would seem, is that in forgoing information, one may eliminate error. And it looks as if a similar method for resolving the trade-off problem must be found. Not so, says Levi. The inquirer is committed to treating the full beliefs collected in his or her corpus as infallible, that is, as standards of serious possibility. From the inquirer's own point of view, his or her beliefs *cannot be* erroneous. Thus, since there is no error that could be eliminated, nothing positive can be gained from a contraction. Only one factor, namely, loss of information, is to be taken into account. We need informational-value determining probabilities, but we do not need credal probabilities any more. The problem of contraction is unequivocal; there is just one thing to take care of: to minimize the loss of information.[14]

[13] It does not make much sense to say that one may import truth. It looks as if all one might wish to convey by this phrase is already covered by saying that the inquirer gains information.

[14] It might seem, then, that in Levi's picture (1) expansions are much more interesting and sophisticated change operations than contractions, and that (2) this is a reversal of the AGM account where expansions are considered to be trivial as compared with contractions. Both

But if, from the inquirer's point of view, there is nothing to be gained from a contraction, why should any rational being contract his or her beliefs in the first place?[15] Levi's answer to this question is three-fold. First, inquirers need to contract their beliefs if they are caught in an inconsistent corpus. Second, they may wish to give a new hypothesis a hearing and for this reason withdraw their full belief in its negation. Third, they may want to engage in reasoning for the sake of argument.

The formats of Levi's expansion and contraction operations are different. In deliberate or inductive expansion, the task is to find the right hypothesis to add to the current corpus, subject to the internal constraint that the expected cognitive value be maximized. Inductive expansion is an autonomous process, a purely internal affair, as it were. There is no input to the inquirer's belief state from outside.

In contraction, on the other hand, the task the inquirer faces is to discard a particular belief h – and *which* belief to discard is externally determined. As h is given, the task is not to find the right hypothesis to subtract. The options that are being judged with respect to their maximizing the inquirer's cognitive values are target corpora that do not contain h.

So how can we characterize the set of options in a contraction problem? The first answer is that every weakening of K is a potential contraction. More precisely: A *potential contraction of K* relative to the ultimate partition U (of the space of all possibilities) is the disjunction of K with the disjunction of some elements of U outside K.[16] A *potential contraction of K removing h* (relative to U) is obtained by forming the disjunction of K with a subset of U outside K that contains at least one element that does not imply h.[17] These are very general concepts of contraction. I agree with Levi that we should not from the outset restrict the possible options for the process of contracting belief sets.

parts are wrong. Statement (2) is superficial because it neglects the fact that AGM simply do not have an operation of inductive expansion (without any input). Statement (1) passes over the fact that finding unique solutions with the help of a "rule of ties" presents a much greater challenge for contractions than it does for expansions. We get back to this point soon.

[15] From a third person's point of view, the answer is trivial: Because inquirers are sometimes mistaken.

[16] Levi (2004, p. 59). If we stuck to the representation of K and the elements of U as *sets of sentences* rather than sentences, then we would have to mention intersections rather than disjunctions.

[17] If such an element exists; otherwise the only potential contraction of K removing h is set to be K itself (Levi 2004, p. 59). I have changed Levi's definition slightly. Where Levi requires that $\sim h$ be implied, I only require that h be not implied. My formulation leaves room for hypotheses that are not (yet) expressible in the inquirer's question-and-answer system. But this is a point that we can neglect in the present chapter.

Levi argues forcefully that hardly any previous author in epistemology has given anything like a decision-theoretic rationale for changes of belief. The following rule for belief contraction is a direct application of this core idea of Levi's:

Rule 1. The Decision-Theoretic Rule

The corpus after a contraction must be optimal, that is, it must minimize the loss of informational value among all corpora expelling the hypothesis *h*.

Levi urges that information (the ruling out of some logical possibilities) must not be identified with informational value. The Decision-Theoretic Rule does not stigmatize every loss of information as irrational. There is no objection to losing worthless information.

The problem is that this rule is not definite enough. There may be many ways to achieve minimum loss of informational value. What is the inquirer supposed to do in the face of such an ambiguity? If two or more options are tied for optimality with respect to the primary decision-theoretic value commitments, Levi recommends invoking a secondary standard of evaluation in order to break the ties:

Rule 2. The Rule for Ties in Contractions

Given a set of optimal contraction strategies, one should always choose the weakest of them if it exists. (2004, p. 119).

This advice is analogous to the Rule for Ties in Expansions. In contrast to the case of expansions, however, the precondition that there is a unique weakest solution is not readily satisfied in the Rule for Ties for Contractions. There is the danger that in many cases the rule simply cannot be applied. Levi, however, feels strongly that there *should always* be a unique weakest optimal contraction strategy. Combine this desideratum with the fact that the most obvious (and perhaps the only principled) solution to the problem of multiple optima is to settle for the combination – intersection of corpora or disjunction of sentences – of all optimal contractions. Then we understand the rationale for Levi's installing a third central condition:

Rule 3. The Intersection Equality

If members of a set *S* of contractions from *K* are equal in informational value, their intersection is equal in informational value to the informational value of any element of *S*.[18]

[18] This is the strong version of Levi's Intersection Equality (2004, p. 125). The weak version restricts the claim to saturatable contractions removing *h* from *K*. As I don't see any reason

Unfortunately, the straightforward way of measuring information does not prove to be suitable for Levi's purposes. Assume that M is an informational-value determining probability function over the class of *all* possibilities (not just those within K).[19] Then the loss of informational value by a shift from K to a potential contraction K' can be determined by $Cont(K) - Cont(K') = (1 - M(K)) - (1 - M(K')) = M(K') - M(K)$ (2004, pp. 84–85, 109). But since the probability of an intersection (or disjunction) of possibilities is in general *more* probable than the possibilities themselves, *more* informational value will be lost by taking intersections (disjunctions) than by taking single possibilities, in violation of the Intersection Equality.

As far as I know, no author except Levi himself has cared to take seriously the ideal of decision-theoretic optimality while at the same time respecting the Rule for Ties.[20] We now turn to Levi's techniques of the damping of informational value that are devised precisely to overcome the tension between the Decision-Theoretic Rule and a nontrivial application of the Rule for Ties.

6. DAMPING INFORMATIONAL VALUE

Having established that purely probability-based informational value ought not to be minimized, Levi suggested two different ways of "damping" informational value in such a way that his desiderata for belief contractions can all be simultaneously satisfied (1991, pp. 127–31; 1996, pp. 262–7; 2004, pp. 125–47). Since there are no analogous problems of multiple solutions for inductive expansions, damping is not necessary there.

6.1 Saturatable sets and damping version 1

Levi's first attempt at solving the problem of contraction centers around the notion of a saturatable set. A *saturatable contraction* of a corpus K removing a hypothesis h is the disjunction (intersection) of K with some (possibly empty) set of elements of U outside K that entail h and a single element of U that entails $\sim h$ (cf. 1991, p. 121; 1996, pp. 20–3; 2004, p. 60). Levi points out that every potential contraction of K removing h can be represented as the disjunction (intersection) of a subset of saturatable contractions of K removing h. This is Levi's *Potential Contraction Condition* (2004, p. 61). This

to bestow a special status on saturatable contractions (see below), I think it is the strong version that captures the essence of the Intersection Equality.

19 The critical questions about the inquirer's probability function M_K raised at the end of section 4 transfer and are indeed aggravated for the impersonal probability function M.

20 But compare the careful discussion in Sandqvist (2000).

is certainly correct, but it does not suffice to accord saturatable contractions removing h an epistemologically distinguished role, since many other types of sets (e.g., maxichoice contractions, saturatable contractions removing $\sim h$) could be mentioned in the Potential Contraction Condition.

Damped informational value version 1 is then defined in two steps (2004, p. 131).[21] The loss of damped informational value incurred by shifting from the corpus K to a saturatable contraction K' removing h is equal to the loss of undamped informational value, that is, $Cont(K) - Cont(K') = M(K') - M(K)$. The loss of informational value incurred by shifting from the corpus K to a disjunction (intersection) of some set of saturatable contractions removing h is then *defined* to be the largest loss incurred by any element in that set, that is, $Cont(K) - Cont(\vee K_i) = Cont(K) - \max_i Cont(K_i) = \max_i M(K_i) - M(K)$.

Levi chose this definition in order to make sure that the informational value of the disjunction (intersection) of two contractions removing h is equal to the minimum of their informational values. Now what Levi recommends (or, rather, recommended) comes down to saying that *the right contraction* removing h from K is the disjunction (intersection) of all saturatable contractions K' removing h that minimize the loss of damped informational value, that is, for which $M(K')$ is minimal (1991, p. 130; 1996, p. 263; 2004, pp. 132–3).[22] It follows from the definition of damping version 1 that any disjunction (intersection) of contractions minimizing loss of informational value minimizes loss of informational value itself. So the Rule for Ties for contractions does not run counter to the Decision-Theoretic Rule.

Unfortunately, the definition as it stands is not well defined. We show this by giving an example. Consider the eight-cell partition U that is generated by the truth-value combinations of the three atomic sentences p, q, and r. Suppose that M assigns probability 0.2 to the two cells with p and q being false, and probability 0.1 to the six remaining cells. Let K be $p \wedge q \wedge r$ (we identify a theory with its generating sentence), and suppose we want to contract K by p. Consider the potential contraction $K' = p \vee q$. Clearly, K' is not saturatable as a contraction that removes p. But K' can be represented as the disjunction $K_1 \vee K_2$ of two saturatable contractions K_1 and K_2 removing p, where $K_1 = p \vee (q \wedge r)$ and $K_2 = p \vee (q \wedge \sim r)$. Since the M-value of both K_1 and K_2 is 0.5, their informational content $Cont$ is 0.5. So by damping version 1, the content of K' is 0.5, too. However, K' can also be represented as the disjunction $K_3 \vee K_4$ of the saturatable contractions K_3 and K_4 removing p,

[21] A similar definition is given in one step in Levi (1996, p. 23).

[22] These are the ones for which the $\sim h$-cell is minimally M-probable among the $\sim h$-cells outside K, and for which the h-cells outside K bear zero M-value.

where $K_3 = (p \wedge \sim r) \vee (q \wedge r)$ and $K_4 = (p \wedge r) \vee (q \wedge \sim r)$. The M-value of both K_3 and K_4 is 0.4, so their informational content *Cont* is 0.6. So by damping version 1, the content of K' must be 0.6, too. But this contradicts the value we have calculated before.

Usually, this ill-definedness does not do any harm, since one has to look at maxichoice (rather than just saturatable) contractions anyway. Levi's first damping construction recommends as the right contraction of K removing p the disjunction (intersection) of the saturatable contractions removing p that minimize loss of damped informational value version 1. In our example, this is $K \dot{-} p = (q \wedge r) \vee (q \wedge (p \leftrightarrow r)) = q \wedge (\sim p \vee r)$, the damped informational value version 1 of which is $\min(Cont(q \wedge r), Cont(q \wedge (p \leftrightarrow r))) = 0.8$. Saturatable sets that are not maxichoice come into play only if there are cells of zero M-value outside K that entail the hypothesis to be removed.

Another point of criticism is more substantial. Damping version 1 obviously bestows a privileged status onto saturatable subsets of K – these are the only sets for which, in Levi's terminology, damped equals undamped informational value. For such a privileged status I can see no good reason. Advocating doxastic conservatism, AGM had begun by considering maximal-nonimplying subsets of K ("maxichoice contractions") as the only options for contraction. Levi is correct in emphasizing that this restriction cannot be justified. As he points out, agents may sometimes turn to logically weaker belief sets without incurring any additional loss in informational value (if the relevant additional possibilities bear zero M-value). However, if this line of reasoning is right, there is no motivation any more for insisting on the property of saturatability that is a remnant of strict conservatism. In sum, I think that Levi's (2004, pp. 134–5) recent decision to give up his earlier "version 1" damping of informational value was a good one.

6.2 Damping version 2

Levi's new theory centers around the old AGM notion of maxichoice contractions. A *maxichoice contraction* of a corpus K is the disjunction (alternatively, intersection) of K with a single element of U outside K (cf. 1996, pp. 20, 262; 2004, p. 60).

Damped informational value version 2 can also be defined in two steps (2004, p. 141). The loss of informational value incurred by shifting from the corpus K to a maxichoice contraction K' is equal to the loss of undamped informational value, that is, $Cont(K) - Cont(K') = M(K') - M(K)$. The loss of informational value incurred by shifting from the corpus K to an disjunction (intersection) of some set of maxichoice contractions removing h is then *defined* to be the largest loss incurred by any element in that set, that is, $Cont(K) - Cont(K_i) = \max_i M(K_i) - M(K)$. Since maxichoice contractions are more definite than saturatable contractions, no problem of well-definedness arises here.

This definition makes sure that the informational value of the disjunction (intersection) of two contractions removing h is equal to the minimum of their informational values. What Levi now recommends comes down to saying that *the right contraction* removing h from K is the disjunction (intersection) of all maxichoice contractions K' that are at least as informationally valuable as a maxichoice contraction removing h that minimizes the loss of damped informational value version 2. This results in the disjunction of K with all the lowest M-valued $\sim h$-cells of U and with all h-cells outside K carrying no higher M-value than these (2004, pp. 142–3). By the definition of damping version 2, this disjunction (intersection) of maxichoice contractions is exactly as informationally valuable as any optimal maxichoice contraction removing h. And it is clearly the weakest one among the maximally informative contractions removing h. Thus there is no problem to apply the Rule for Ties for contractions, and to combine it with the Decision-Theoretic Rule. Levi calls this method of removing a hypothesis from a corpus *mild contraction*.

I like Levi's idea of refocusing on AGM's maxichoice contractions rather than saturatable contractions, but I find his concept of damped informational value version 2 counterintuitive.

To see this, let us first have a fresh look at the example of the previous section. The maxichoice contractions removing p with the smallest loss of information are $K \vee (\sim p \wedge q \wedge r) = q \wedge r$ and $K \vee (\sim p \wedge q \wedge \sim r) = q \wedge (p \leftrightarrow r)$, each bearing M-value 0.2 and $Cont$-value 0.8. But all the maxichoice contractions that do *not* remove p have the same values. So what Levi recommends is actually the disjunction (intersection) of six maxichoice contractions with these values. It is easily seen that this disjunction is $K' = p \vee q$, the damped informational value version 2 of which is the minimum of the $Cont$-values of the maxichoice contractions involved, that is, 0.8.

We have finally found that according to damping version 2 the belief state represented by $K' = p \vee q$ has the same informational value as any one of the six maxichoice contractions that are used for the construction of K'. This is strange, since K' is obviously *much* weaker than the latter contractions (it comprises six as opposed to only two cells of the ultimate partition). What could be the justification for this deviation from our ordinary intuition of informational value? I cannot think of one.

More generally, a typical situation is this. Suppose K is a corpus and K_1 and K_2 are two different proper subsets of K that both minimize the loss of information, subject to the constraint that h be removed from K. Let us suppose that K_i is the disjunction (intersection) of K and a single cell x_i of U outside K ($i = 1, 2$), with x_1 different from x_2. That is, K_1 and K_2 are maxichoice relative to U. Suppose further that both K_1 and K_2 incur a nonzero loss of informational value. This means that if the inquirer starts from the corpus K, admitting the possibility x_1 loses some informational value, and admitting the possibility x_2 loses some informational value as well. My thesis now is that on any natural account of informational value, admitting the possibility x_1 also incurs a nonzero loss of information if the inquirer were to start from K_2, and admitting the possibility x_2 also incurs a nonzero loss of information if the inquirer were to start from K_1. The *amount* of information lost may vary from context to context, but the fact that *some* information is lost if the inquirer ceases to be able to rule out a possibility seems indisputable, given that this fact means a loss of informational value when the inquirer sets out from K. The particular corpus from which the inquirer starts should not make *that* much of a difference. But if this is right, it follows that $K_1 \vee K_2$, which rules out neither x_1 nor x_2 has *less* informational value than either of K_1 and K_2, and hence its informational value is *lower* than the minimum of the values of K_1 and K_2.

This consideration is very close to the Principle of Constant Marginal Returns that Levi endorses in the context of deliberate expansions (see section 3 above). Unfortunately, he does not comment on why he refrains from employing the same principle for belief contraction.

Levi (2004, pp. 181–6) points out that his model of mild contraction is formally identical with a model studied under the name *"severe withdrawal"* by Pagnucco and Rott (1999). In that paper, the model was justified in terms of Principles of Preference and Indifference, and this still seems more convincing to me than Levi's justification in terms of damped informational value version 2. Levi is exactly right in saying that we all (Levi, Pagnucco, and Rott) agree that mild contraction *alias* severe withdrawal should be "taken seriously" (2004, p. 147). But I personally think that neither Levi's nor Pagnucco's and my justification of mild contraction is strong enough to warrant its endorsement as *the* distinguished legitimate way of contracting corpora of belief.[23] Levi's mild contraction surely deserves to be taken seriously, but it is very severe indeed, much more severe than AGM

[23] Hansson's (1999, observation 2.52) criticism that mild contraction is too "expulsive" still stands unanswered: For any two hypotheses h and h', the inquirer either loses h' when removing h or she loses h when removing h' from her corpus – even if the contents of h and h' are in no way related.

contraction and his own contraction based on damped informational value version 1.

In the past couple of years, Levi has come to advocate strongly version 2 of damped informational value. I think that this notion is motivated by his wish to construct a measure of information that conforms both to the Decision-Theoretic Rule and the Rule for Ties at the same time.[24] I am not convinced that this project is on the right track. I rather think one should acknowledge that the desiderata expressed by the two rules tend to require genuinely different kinds of contraction behavior: The former requires informational economy (or "thrift"), the latter requires the equal consideration of multiple solutions (or "fairness").

Pagnucco and Rott (1999) argued that the AGM model of partial meet contraction is above all committed to the Rule for Ties, since partial meet contraction effectively makes considerations of fairness override considerations of minimal change (or maximum information). The strict idea of minimal change would indeed insist on maxichoice contraction. For AGM, the Rule for Ties is not, as Levi's picture suggests, a secondary value commitment that comes after the idea of minimizing loss of informational value (Levi 2004, pp. 119, 150–1). It is a primary value commitment. This is not in conflict with Levi's own ideas, since, as we saw, he claims that we can fully satisfy the idea of fairness, provided that we use the right notion of informational value: damped informational value version 2. But he goes farther than AGM in claiming that he can at the same time be true to the idea of minimizing the loss of informational economy.

To me it seems that the *only* support for Levi's notion of damped informational value version 2 is that it represents a way of simultaneously satisfying the two desiderata of informational economy and fairness. But the whole strategy itself seems counterintuitive to me. Rule 1 and Rule 2 give expression to conflicting desiderata, and I see no reason why one should not say so. As we all know, economy and fairness just don't always agree. Agents have to rank or to weigh them in order to reach principled and satisfactory decisions.[25] It is somewhat ironic that Levi, who has said so many insightful things about

[24] See Levi (2004, p. 125): "When two or more saturatable contractions removing *h* minimize loss of informational value, we *want to* be in a position to recommend adopting the 'skeptical' contraction that is the intersection of these optimal contractions. And we *want to* be in a position to do so while still claiming that informational value is being minimized" (my emphasis).

[25] Many people will be inclined to think that the pair of payoffs $\langle 60, 40\rangle$ is "better" than both of the payoff pairs $\langle 45, 45\rangle$ and $\langle 80, 30\rangle$. The latter is too unjust, the former too wasteful.

Hard Choices, has here refused to acknowledge the existence of conflicting desiderata in belief contraction.

7. CONCLUSION

I have tried to present Levi's basic decision-theoretic models of inquiry as perspicuously as possible. His account of deliberate or inductive expansion (without inputs) combines the concerns for truth and information in a very interesting and elegant manner, successfully linking inductive reasoning to decision theory. On the other hand, I find Levi's account of belief contraction (with regard to some preselected hypothesis) not quite as convincing. His aim is to obey two rules for belief contraction at the same time: the Decision-Theoretic Rule of minimizing the loss of informational value and the Rule for Ties. To achieve this aim, he has invented two notions of damped informational value that satisfy the Intersection Equality. I have argued that both notions lack independent motivation and thus fail to render compatible the two desiderata that Levi set out to meet. These desiderata just pull in different directions. It would be nice if Levi were right, but I think we have to put aside our aspirations toward theoretical elegance and admit that a compromise between (i.e., intersection or disjunction of) optimal solutions need not itself be optimal in the same sense. Obeying the tie-breaking rule means sacrificing some informational value.

For reasons of space, I have refrained from making an issue of another, more fundamental point. Levi puts great emphasis on an important asymmetry between belief expansion and belief contraction. The notion of truth – or, more precisely, of probability of truth – is crucial in belief expansion but is completely absent in belief contraction. The reason is that Levi's inquirer is conscious of the possibility of importing error through expansion, but completely rules out the possibility of having an error contained in his or her current corpus of full beliefs:

> If an inquirer's belief state is K, any hypothesis h incompatible with K is not a serious possibility. The coherent inquirer must regard every such hypothesis as certainly false *and maximally implausible*.... [A]s long as long as the inquirer is in the state of full belief represented by K, the inquirer cannot coherently acknowledge the serious possibility of being mistaken in this belief.... Each and every cell in U^*_K [in U outside K] is maximally and equally implausible. There can be no distinction between hypotheses incompatible with K with respect to plausibility. (Levi 2004, p. 174)

This dogmatic insistence of the inquirer that all of his or her current beliefs must be true is a central feature of Levi's brand of pragmatism. In his terminology, inquirers are not *incorrigible* (otherwise, there would be no contractions), but they are bound to consider themselves *infallible*. All beliefs in the corpus K have maximal degree of belief.[26] This is why credal probabilities need not be considered in Levi's contraction, and only informational-value determining probabilities matter.

My feeling is that this (temporary) dogmatism of the inquirer about her beliefs might overtax pragmatism to a point where it becomes flatly implausible. First, it is one of Levi's ideas that contractions are performed in order to give alternative hypotheses a hearing. Why should an inquirer do this if she is certain that she is right anyway? She must be having a faint idea that there might be something wrong with her beliefs. Second, at the time an inquirer incorporates new beliefs into her corpus by expansion, she must be aware that inductively expanding her beliefs is risky business. Why should she be committed to forgetting all about the previously uncertain status of her newly promoted full beliefs? Insofar as these questions are to the point, inquirers are and should be concerned about truth not only in expansions, but also in contractions. Error cannot, of course, be imported in contractions, but it can be expelled by contractions. The challenge is to come up with a model that reverses Levi's model for expansions and uses sensible, well-motivated credal and informational-value determining probability functions.

Usual notions of justification play no role in Levi's epistemology, which thus seems orthogonal to the traditional concerns of mainstream theories of knowledge.[27] Levi's pragmatist attitude opposes pedigree epistemology, holds that it is not beliefs but changes of belief that are in need of justification, and gives such justification in decision-theoretic terms. At the time of writing, Levi's immensely important approach is still only loosely connected with mainstream epistemology. This is a regrettable state of affairs, and one I hope will be alleviated soon.

[26] According to Shackle's way of determining degrees of *belief*. For those whom Levi calls "Parmenidean epistemologists," including Gärdenfors and Spohn, only logical truths have maximal degree of *invulnerability*. In essence, Shackle belief functions as used by Levi order nonbeliefs, while invulnerability functions order beliefs. Shackle functions are useful for inductively expanding K into some decision-theoretically recommended $K + h$, invulnerability functions are useful for contracting K into some decision-theoretically recommended $K \dot{-} h$. Shackle measures are dependent on K, U, Q_K, and M_K, while invulnerability measures are dependent on K, U, M, and, possibly, on the sentence to be retracted (Levi 1996, p. 267; 2004, pp. 192, 196).

[27] A still inspiring intermediary can be found in Harman (1986).

Carnap, Rudolf, and Yehoshua Bar-Hillel. 1952. "An Outline of a Theory of Semantic Information." Technical Report 247, Research Laboratory of Electronics MIT. Reprinted in Yehoshua Bar-Hillel, *Language and Information*, pp. 221–74. Reading, Mass., and London: Addison-Wesley, 1964.

Fioretti, Guido. 2001. "A Mathematical Theory of Evidence for G. L. S. Shackle." *Mind and Society: A Journal of Cognitive and Epistemological Studies on Economics and Social Sciences* 2: 77–98. Http://www.icer.it/docs/wp2001/Fioretti3-01.pdf.

Hansson, Sven Ove. 1999. *A Textbook of Belief Dynamics: Theory Change and Database Updating*. Dordrecht: Kluwer.

Harman, Gilbert. 1986. *Change in View*. Cambridge, Mass.: Bradford Books and MIT Press.

James, William. 1979. "The Will to Believe." In F. H. Burkhardt, F. Bowers, and I. K. Skrupskelis (eds.), *The Will to Believe and Other Essays in Popular Philosophy – The Works of William James*, vol. 6, pp. 13–33. Cambridge, Mass., and London: Harvard University Press. Originally published in 1897.

Kaplan, Mark. 2000. "Decision Theory and Epistemology." In Paul K. Moser (ed.), *The Oxford Handbook of Epistemology*, pp. 434–62. Oxford: Oxford University Press.

Levi, Isaac. 1967. *Gambling with Truth: An Essay on Induction and the Aims of Science*. New York: Alfred Knopf.

Levi, Isaac. 1969. "Information and Inference." *Synthese* 17: 369–91. Reprinted in Isaac Levi, *Decisions and Revisions* (Cambridge: Cambridge University Press, 1984), pp. 51–69.

Levi, Isaac. 1980. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. Cambridge, Mass.: MIT Press.

Levi, Isaac. 1991. *The Fixation of Belief and Its Undoing: Changing Beliefs through Inquiry*. Cambridge: Cambridge University Press.

Levi, Isaac. 1996. *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference and Nonmonotonic Reasoning*. Cambridge: Cambridge University Press.

Levi, Isaac. 2002. "Seeking Truth." In Wolfram Hinzen and Hans Rott (eds.), *Belief and Meaning: Essays at the Interface*, pp. 119–38. Frankfurt: Hänsel-Hohenhausen.

Levi, Isaac. 2004. *Mild Contraction: Evaluating Loss of Information Due to Loss of Belief*. Oxford: Oxford University Press.

Pagnucco, Maurice, and Hans Rott. 1999. "Severe Withdrawal (and Recovery)." *Journal of Philosophical Logic* 28: 501–47.

Sandqvist, Tor. 2000. "On Why the Best Should Always Meet." *Economics and Philosophy* 16: 287–313.

Shackle, George L. S. 1949. *Expectation in Economics*. Second edition, 1952. Cambridge: Cambridge University Press.

Shafer, Glenn. 1976. *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

14

Decision-Theoretic Contraction and Sequential Change

Horacio Arló Costa

For almost half a century Isaac Levi has perfected the details of a comprehensive epistemology deeply rooted in the epistemic voluntarism of the three great American pragmatist philosophers (J. Dewey, C. S. Peirce, and W. James). Beliefs in his view "are not impressions or dispositions to assent or other behavior, they are commitments" (Levi 1991, p. 74). Levi would argue that rational agents are able to choose how to revise these doxastic commitments and that, in this sense, agents can deliberately change their beliefs. So, rational agents can choose what to believe. Not always, of course. Perception, for example, would be in Levi's view a case of *routine expansion* where the inquirer expands his or her view in accordance with a precompiled program for adding new information. But forming a belief via one of these routines involves in Levi's view forming a commitment just as much as when the inquirer deliberately chooses to be committed.

Levi's pioneering work on belief revision was first presented in his important book *The Enterprise of Knowledge* (Levi 1980), building on work on induction first presented in 1967. The mid- and late 1980s witnessed the beginning of a multidisciplinary discussion on the nature of belief change inaugurated by the publication (in 1985) of an influential paper by Carlos Alchourrón, Peter Gärdenfors, and David Makinson. The AGM trio was of course aware of Levi's work but they tackled the problem of belief change under a very different point of view. They focused on axiomatizing a belief change operation, and the algebraic semantics they deployed for their axioms

This chapter was written enjoying a sabbatical leave from my duties at Carnegie Mellon. During the spring semester of 2004 I presented earlier drafts of this chapter both at a meeting of the group on Belief Revision and Decision Theory at Columbia University and at a weekly meeting of the seminar on Logic and Games (organized by Rohit Parikh at CUNY Graduate Center). I would like to thank all the participants of these groups for helpful comments. I am especially grateful to Issac Levi for insightful advice and for a number of valuable suggestions.

abstracted away from most of the decision-theoretical problems that were Levi's center of attention. Moreover, some of the members of the AGM trio were convinced Popperians with little or no sympathy for induction, a problem that for Levi was central since at least 1967. This is reflected, for example, by the fact that expansion in the AGM theory is the only unproblematic change operation reduced to just taking the logical closure of the background theory with the given input; while for Levi expansion is one of the central change operations instantiating a case of ampliative or inductive inference.

Since the late 1980s to the present day (and especially during the 1990s) the community of researchers in artificial intelligence (especially the logical community dealing with knowledge representation) noticed that the topic of belief change was central for their research agenda, and since then a barrage of literature was produced in the area. This work followed mostly the trend initiated by the AGM paper, and therefore it was for the most part unconnected with to the decision-theoretic problems associated with changing view (or at most attentive to the sort of sophisticated probabilism influencing the work of Wolfgang Spohn (1988)).

Levi responded to some of this literature, especially to the AGM work, first in a book published in 1991, then in another book published in 1996 (centering more specifically on issues related to computer science). And finally, he has just submitted the manuscript of a third book, which is also part of an ongoing attempt to formulate a decision-theoretically motivated theory of contraction. The new book is in part a rejection of a theory of contraction proposed in Levi's 1991 monograph. So this recent work is part of a continuous effort to perfect his account of belief change. But part of it is also polemic. In fact, he has tried to argue in most of these works that there is much to lose by ignoring the fact that a deliberate change of view ultimately involves a rational choice. He has therefore argued repeatedly for the view that changes in belief ought to be evaluated according to principles of rational choice relative to appropriate cognitive goals.

I think that Levi is right concerning the last important point. It is also increasingly clear that the axioms that completely characterize a decision-theoretically motivated notion of contraction do diverge from the ones initially adopted by AGM. Levi and I recently presented a proof (Arló-Costa and Levi 2004) of a complete syntactic representation of one of the central notions of contraction studied in Levi (2004) where this divergence is clear. The notion of contraction axiomatized in Arló-Costa and Levi (2004) fails to obey some AGM axioms (such as Recovery), and it obeys postulates (such as Antitony) that fail to be satisfied by AGM (and many of its studied alternatives). The systematic study of the logical properties of decision-theoretically motivated notions of contraction and revision is only incipient. Nevertheless, there are many interesting and unexplored problems in this area whose detailed study could offer a better picture of the connections of the decision-theoretical approach with rival approaches. In particular, I try to mend bridges here between the decision-theoretic perspective and other contemporary work in belief change done mostly by computer scientists. One of the focuses of the paper is the extension of the theory presented in Arló-Costa and Levi (2004) to cover iterated change (in particular, iterated supposition as preliminarily presented in Levi 1996). I also offer a preliminary syntactic account of some notions of contraction, which can be parametrically accommodated by adopting corresponding constraints on the underlying value function. This chapter focuses on sketching possible directions for research in this area. Of course, attractive (and so far unexplored) avenues of inquiry are not even considered here due to space limitations. One salient example is the family of philosophical and technical issues related to extending the decision-theoretical approach of belief change to the case where information value is indeterminate or vague (this issue is considered in Arló-Costa 2005). My main goal is to persuade the reader of the interest and fertility of Levi's account of belief change by showing its generality and its capacity for solving open problems not only in this area but in many other epistemological applications. Levi's epistemology is a fascinating alternative to some of the established views in the field. It has important consequences for many areas of philosophical logic, artificial intelligence, and decision theory, just to mention three relevant fields. These communities have much to gain, I would like to argue, by further articulating and applying the epistemological theories that Levi pioneered and developed throughout his career.

1. INFORMATIONAL VALUE AND CONTRACTION

Let L be a classical propositional language containing the classical connectives. The underlying logic will be identified with its Tarskian consequence operator Cn: $2^{L} \rightarrow 2^{L}$. We also assume that Cn obeys the deduction theorem and is compact. A *theory* is any set K such that $K = Cn(K)$. Theories can be used advantageously in order to represent the epistemic commitments of rational agents.

According to Levi, in giving an account of belief change, it is desirable to focus on changes in theories *relevant* to a given problem or question or cluster of questions. The idea here is that a problem or question in inquiry typically presupposes substantial claims that are intended to be taken for granted throughout the changes in belief that take place. Levi proposes to gather these assumptions in a minimal theory LK included in the current view K.

The potential answers to the problem under consideration are then arranged into a basic partition $\mathcal{B}$ where each cell in this partition is an expansion of the theory LK. A necessary constraint on the admissibility of $\mathcal{B}$ is that should be formed by expanding LK with sentences that are relevant answers to questions under investigation and that the expansions are restricted to expansions by adding to LK elements of a set of sentences such that LK entails that exactly one of them is true and each element of the set is consistent with LK. The subpartition of $\mathcal{B}$ constituted by the partition cells whose intersection is K will be called $\mathcal{U}$ and its complement $\mathcal{D}$. Levi for the most part restricts his discussion to the finite case ($\mathcal{B}$ is finite). I adopt the same constraint here and I assume as well that all partition cells are finitely axiomatizable. Of course $\mathcal{B}$, $\mathcal{U}$, and $\mathcal{D}$ are all relative to K and to LK. I shall omit subindexes here for the sake of readability.

Each cell in the basic partition can be represented as the intersection of a family of maximal and consistent sets of the initial language L. I adopt the following notation. $|A|$, for $A \in L$, denotes the set of maximal and consistent extensions M of L such that $A \in M$. For any theory K such that the theory can be represented as a the intersection of a set of partition cells $C_1, \ldots, C_n$, I use the notation [K] to denote $\{C_1, \ldots, C_n\}$. Also, if A is the finite axiomatization of K, $[A] = \{C_1, \ldots, C_n\}$. In general $[A] = \{C_i \in \mathcal{B}\colon A \in C_i\}$. $L \supseteq \mathbf{L} = \{A \in L\colon [A] \neq \emptyset$, and $[\neg A] \neq \emptyset\}$.

We can immediately define some useful notions. Every *potential contraction removing $A \in \mathbf{L}$* from K is the intersection with K of a nonempty subset R of $\neg A$-entailing cells of $\mathcal{D}$ and a subset R^* of A-entailing cells of $\mathcal{D}$ that may or may not be empty. A *maxichoice contraction* of K relative to $\mathcal{D}$ is the intersection of K with a single element of $\mathcal{D}$. A *maxichoice contraction of K removing $A \in \mathbf{L}$* relative to $\mathcal{D}$ is the intersection of K with a single element of $\mathcal{D}$ that entails $\neg$A. A *saturatable contraction* of K *removing $A \in \mathbf{L}$* relative to $\mathcal{D}$ is the intersection of a maxichoice contraction of K removing A relative to $\mathcal{D}$ with the intersection of a set of elements of $\mathcal{B}$ none of which entail A.

DEFINITION. Let S(K;A) be the family of A-saturatable sets of K; that is, if K is a theory, $X \in S(K;A)$ if and only if $X \subseteq K$, X is closed, and $Cn(X \cup \{\neg A\})$ is an element of the partition $\mathcal{D}$.

$\Theta = \{X\colon X = \cap Y$, with $Y \in 2^{\mathcal{D}} \cup [K]\}$. Now we can now introduce a *measure of informational value* V: $\Theta \rightarrow \mathbf{N}$ (natural numbers). As the terminology indicates, V is supposed to deliver a measure of the value of *information*. As such, Levi assumes that it inherits some basic properties of classical measures of information that are probability-based. A classical manner of utilizing probability in order to measure the content of information is to utilize the measure Cont(.) = 1 – Prob(.). There are two basic properties that probability-based measures of information satisfy. First, they *respect entailment* in the following sense:

(*Weak Monotonicity*) For any two sets X, Y $\subseteq \Theta$ such that X $\subset$ Y, V(X) $\leq$ V(Y).

The second important postulate is the following one:

(*Extended Weak Monotonicity*) Let X;Y $\subseteq \Theta$. If S $\subseteq \Theta$ is incompatible with both X and Y, and if V(X) $\leq$ V(Y), then V(X $\cap$ S) $\leq$ V(Y $\cap$ S).

Unfortunately, one cannot preserve all the properties of Cont in characterizing a notion of information value useful in contraction. The trouble with Cont is that it cannot rationalize (in terms of optimality) moving to a position of suspense when there is a tie in optimality. In fact, the Cont-value of the intersection of two optimal saturatable contractions need not and, in general, will not carry maximum Cont-value. So Levi proposes to preserve the first two postulates while adding a third that permits rationalizing suspense among optimal options as optimal. To present this third postulate, we need an additional piece of notation. Notice first that any saturatable contraction in S(K;A) has the canonical form $K \cap T_A \cap m_{\neg A}$, where T_A is an intersection of A-cells in $\mathcal{D}$ and where $m_{\neg A}$ is a ¬A-cell in $\mathcal{D}$. Then we can say that two saturatable contractions removing A from K are *A-equivalent* if and only if they are constituted as intersections of K with different ¬A-cells in $\mathcal{D}$ and the same subset T_A of the subset all of whose members entail A. A saturatable contraction S removing A is A-equivalent to an intersection of a set T of saturatable contractions removing A (including S) if S is constituted as the intersection of K, a set T_A of A-entailing cells and a ¬A-cell in $\mathcal{D}$, and $[\cap T \cap A] = T_A$.

(*Weak Intersection Equality*) For every subset *T* of potential contractions removing A from K each element of which is of equal informational value and such that all the elements in *T* are A-equivalent and A-equivalent to their intersection, then for every X $\in$ T, V($\cap$T) = V(X).[1]

[1] This is, I understand, the most recent version of this postulate held by Levi (personal communication). Previous versions did not appeal to constraints in terms of *A*-equivalence. In particular, it is useful to see that the following postulate *does not* follow from the one stated above:

> For every subset *T* of *S*(*K*, *A*) each element of which is of equal informational value and for every $X \in T$, $V(\cap T) = V(X)$.

These three *core* postulates jointly imply (Arló-Costa and Levi 2004):

(*Weak Min*) If T is a finite subset of S(K,A), $V(\cap T) = \min(V(X): X \in T)$.

Some of the core conditions (WM especially) induce important constraints on contraction.

DEFINITION $\div$ is an operator of informational value for a closed set K if and only if there is a selection function γ such that for all A in **L**: (i) if $A \in K$, then $K \div A = \cap\gamma(S(K;A))$, where $\gamma(S(K;A)) = \{X \in S(K;A): V(Y) \leq V(X)$ for all $Y \in S(K;A)\}$ and (ii) $K \div A = Cn(K)$ otherwise.

Observation 1 (Hansson and Olsson 1995). A *basic* operator of informational value (obeying Weak Monotonicity) obeys the following conditions ($\div$1) to ($\div$ 6).

($\div$1) $K \div A = Cn(K \div A)$ [closure]

($\div$2) $K \div A \subseteq K$ [inclusion]

($\div$3) If $A \notin K$ or $A \in Cn(LK)$, then $K \subseteq K \div A$ [vacuity]

($\div$4) If $A \notin Cn(LK)$, then $A \notin K \div A$ [success]

($\div$5) If $Cn(A) = Cn(B)$, then $K \div A = K \div B$ [extensionality]

($\div$6) If $A \notin K \div (A \wedge B)$, then $K \div (A \wedge B) \subseteq K \div A$ [conjunctive inclusion]

It is important to see that the core postulates *do not* validate some stronger syntactic postulates of contraction, which have been discussed at length in the literature. I list here two of these postulates:

($\div$7) If $A \notin Cn(\emptyset)$, then $K \div A \subseteq K \div (A \wedge B)$ [antitony]

($\div$8) If $A \notin Cn(\emptyset)$, then $Cn(K \div A) \cup \{A\}) = K$ [recovery]

Recovery is part of the AGM theory of contraction, but Levi has argued at length against its tenability (Levi 1991). So, he has proposed (2004) that a condition of adequacy of stronger theories of informational value (obtained by imposing further constraints on V aside from the core postulates) is that they should not lead to the validation of recovery. Of course this constraint is rather weak. There is a relatively large spectrum of permissible stronger theories satisfying this constraint. In Arló-Costa and Levi (2004) an argument is presented for selecting exactly one theory among the permissible ones. There is a strengthening of the core postulates leading to a theory of contraction, which can also be rationalized independently in terms of a direct articulation of the notion of entrenchment and its role in contraction. The idea is that when A is given up from a theory K, one should preserve all sentences better entrenched than A. A theory of informational value compatible with this simple idea is the one obtained by requiring in addition to the core postulates:

(*Strong Intersection Equality*) If T is a set of maxichoice contractions from K each element of which is of equal informational value and for every $X \in T$, $V(\cap T) = V(X)$.

Strong intersection equality combined with the core postulates entails the following:

(*Min*) If X and Y are contractions from K, $V(X \cap Y) = min(V(X), V(Y))$.

Arló-Costa and Levi (2004) have shown that the resulting operator of informational value can be completely characterized syntactically in terms of the postulates ($\div$1) to ($\div$6), plus the postulate of Antitony presented above (Levi calls the resulting notion *mild contraction*, while Rott and Pagnucco call it *severe withdrawal*). Antitony is less known than recovery, but so far it has produced an amount of controversy similar to that of recovery. Some scholars strongly oppose it. For example, Hansson has argued (1999) that "Antitony does not hold for any sensible notion of contraction." The following section is devoted to develop models for Antitony as well as reviewing some alternative options for strengthening the basic postulates without imposing neither Recovery nor the full strength of Antitony.

Of course, the three core postulates are constitutive of any permissible notion of information-value as Levi understands it. It would be desirable to have a complete syntactic characterization of the corresponding notion of core contraction. In particular, it would be nice to know whether EWM and WIE induce the validity of contraction axioms not entailed by the basic postulates.[2]

2. SYSTEMS OF SHELLS OF INFORMATIONAL VALUE

The proof of Antitony offered in Arló-Costa and Levi (2004) takes advantage of a representation of standard operators of informational value in terms

[2] There is an different manner of reflecting WIE and EWM syntactically by introducing a binary relation $\leq_V$ with the following interpretation:

$$P \leq_V Q \text{ if and only if } V(P) \leq V(Q), \text{ for } P, Q \subseteq 2^{\mathcal{D}}.$$

This would permit a direct encoding of WIE and EWM. Having a language with this expressive power could be useful for other purposes as well. The question remains open, nevertheless, as to whether the basic postulates are enough to fully characterize core contraction (in a language only containing $\div$) or whether independent axioms (constraining $\div$) weaker than Antitony are needed in order to accomplish a complete representation of core contraction.

of an operation defined in terms of the *system of shells of informational value* (this representation is also crucially used there in order to prove the main completeness result offered in the article). Systems of shells also play a central role in analyzing sequential change below and in many other arguments presented in the following sections.

DEFINITION Let $I = \text{range}(V)$ be a set of indices. For $x \in I$ let R^X be the non-empty set X of cells in $\mathcal{D}$ such that for every $Y \subseteq X$, $V((\cap Y) \cap K) = x$.

Intuitively, R^X groups the $\mathcal{D}$-cells such that the intersection of each of them with K has value x. By Min the intersection of any subset of them with K also has value x. We can extend here the notion of rank, by adjudicating ranks to propositions $P \subseteq 2^{\mathcal{D}}$.

$$\rho^+(P) = \max(y\colon R^y \cap P \neq \emptyset)$$

This notion of rank will be useful below. We can now introduce the notion of m-shell of informational value. The idea of an m-shell is to group together all the ranks R^X where x is greater than or equal to the index m.

DEFINITION The m-shell of informational value $S^m = \cup\{R^x\colon X \in I \text{ and } x \geq m\}$. $\mathcal{S}$ is a system of shells of informational value if $\mathcal{S} = \{S^X\colon \cup S^X = \mathcal{D}\}$.

It is obvious that shells are nested. So a system of shells for a function V determines (at least) a grading of the cells in $\mathcal{D}$. It is important to see here that for any cell $w \in \mathcal{D}$ we do not necessarily have $V(w) = \rho^+(w)$. The value-level of a cell in $\mathcal{D}$ need not coincide with its rank.

With the help of the previous definitions we can now characterize our standard operator of informational value as an operation defined in systems of shells of informational value. We only need an additional definition. Let a sentence A be *rejected* in K if and only if $\neg A \in K$.

DEFINITION Let $A \in \mathbf{L}$ be a sentence rejected in K. Then S_A is the union of [K] with the set $X \in \mathcal{S}$ such that $X \cap [A] \neq \emptyset$ and for any other $Y \in \mathcal{S}$, such that $Y \cap [A] \neq \emptyset$, $X \subseteq Y$.

S_A just picks the union of [K] with the innermost shell in the $\mathcal{S}$ for V containing A-cells. Let's call standard any operator of informational value defined via definition 3 where the underlying value function obeys the core postulates plus Min. Then we can characterize standard operators in terms of the operation S_A defined in terms of systems of shells:

Observation 2. $[K \div \neg A] = S_A$

Standard operators of informational value can be characterized purely in terms if ranks by defining an additional "lower" rank:

$$\rho^{-}(P) = \min(y\text{: } R^{y} \cap P \neq \emptyset)$$

Lower ranks have some properties obeyed by ranking operations in systems such as Spohn's. For example: $\rho^{-}(P \cup Q) = \min(\rho^{-}(P), \rho^{-}(Q))$. Now we have the following corollary:

Corollary 1. $[K \div \neg A] = \cup\{P \in 2^{\mathcal{D}} : \rho^{-}(P) = \rho^{+}(A)\} \cup [K]$

Ranking systems as defined by Spohn and other scholars should be distinguished from shell systems. A detailed comparison stressing formal differences is presented in Arló-Costa and Levi (2004). With the help of the elements just introduced it is easy to show that Antitony holds when the value function is constrained by the core postulates and Min:

Observation 3 (Arló-Costa and Levi 2004). A standard operator of informational value satisfies Antitony.

So, the addition of Min to the core postulates makes Antitony valid, without validating Recovery. Moreover, the completeness result in Arló-Costa and Levi (2004) guarantees that the addition of Antitony to the postulates ($\div$1) to ($\div$6) offers a complete axiomatization of standard operators of informational value. But, of course, the basic postulates ($\div$1) to ($\div$6) can be strengthened by the addition of postulates weaker than Antitony generating contraction theories that also fail to satisfy Recovery. The weakest such theory is the theory of core contraction considered above.

Of course, one can study different kinds of core contraction operators. Let A, E, be consistent and nontautological sentences where both A, E are accepted in K. Then there is a postulate formulated purely in terms of the contraction operator $\div$ that is obeyed by some notions proposed in the literature.

(**Inclusion**) If E entails A and $A \in K$, $E \in K$, then
$[K \div E] \cap [E] \subseteq [K \div A] \cap [A]$.

There are a number of concrete cases that fall under this general case. An example is the case where contraction has the canonical form: $[K/A] = [K] \cup I \cup (R^{x} \cap [\neg A])$, with $\rho^{+}(\neg A) = x$ and where I denotes the innermost shell. But the postulate is rather general. For example, it holds both for partial meet contraction and for mild contractions as well.[3]

[3] The postulate would rule out partial meet contraction in case of requiring that $[K \div E] \cap [E]$ should be nonempty.

The following postulate limits the options permitted by Inclusion by ruling out mild contraction as a possible $\div$ operator:

(**Covering**) If $B \in K = Cn(A)$ and $[\neg B] \cap [K \div A] = \emptyset$, then $K \div B \subseteq K \div A$ if and only if $[K \div B] \cap [B] = [K \div A]$.

According to the classification offered in Appendix B of Rott and Pagnucco (1999), this axiom would also be satisfied by the notion of contraction studied by Levi (1991), which Levi calls in his new book *damped informational value type 1*. The notion in question can be defined in two steps. Each cell receives an M-value (where M is formally a probability function). Then what Levi calls undamped informational value is measured by $\text{Cont}(.) = 1 - M(.)$.

Now the loss of damped informational value incurred by shifting from the corpus K to a saturatable contraction removing A is equal to the loss of undamped informational value, incurred in the change. And the loss of informational value incurred by shifting from the corpus K to an intersection of an A-equivalent set of saturatable contractions removing A is the largest loss incurred by any element in that set.

The smallest rank for an operation of this kind contains the cells receiving the smallest M-value (i.e., the largest Cont-value). In particular, if there are cells receiving zero value, they will constitute the smallest shell (rank). So, we see that in this latter case, Covering has to be satisfied by the type 1 operator of information value. For assume that $B \in K = Cn(A)$, then $[K \div_1 A] = [K] \cup I$, where I is the innermost rank. A is a "covering" sentence that yields the weakest possible contraction, such that $[K \div_1 A] = [K] \cup I \subseteq [K \div_1 B]$, for every sentence $B \in K$. Moreover, $[K \div_1 B] \cap [B]$ will exactly be $[K \div_1 A]$ in this case (where the innermost rank is constituted by the cells of zero M-value). But the M function might assign positive probability to every cell in the partition. In this case Covering continues to be true, given that both the right-hand side and the left-hand side of the biconditional in the consequent of the axiom are false. But the contraction defined above ($[K/A] = [K] \cup I \cup (R^x \cap [\neg A])$, with $\rho^+(\neg A) = x$, where I denotes the innermost shell) also satisfies Covering. And the same holds for partial meet. By appealing to iteration, one can put further constraints that rule out the operator /.

If $Cn(A) = K$ and $Cn(B) = K \div A$, and $(K \div A) \div B \supset Cn(D)$, then $[(K \div A) \div D] \cap [D] = [K \div A]$

If there is an innermost rank I of cells of zero M-value, we have (as before) that $[K/A] = [K] \cup I$. But in calculating contractions from $K \div A$ one has to work with a new shell system where there are no ranks containing cells carrying zero M-value (or, in terms of the partition, the initial dual partition $\mathcal{D}$ is transformed into a new partition $\mathcal{D}'$

that does not contain cells carrying zero M-value). So, $[(K/A)/B] = [K/A] \cup (R^x \cap [\neg B])$, with $\rho^+(\neg A) = x$. So, if I' is the innermost rank of the transformed shell system after contracting A, we have $[(K/A)/B] = [K/A] \cup I'$. And $[(K/A) \div D] \cap [D] = [K/A] \cup I'$ as well. The reader can appreciate that we are calculating all iterated information value operations with respect to an underlying unmodified value function (and M-function). Here we have not elaborated on the details of this extension for iteration. A more detailed analysis of iteration for mild contractions is offered below. As we remarked before in the case of core contraction, it would be interesting to have a complete characterization of operators of damped information type 1. Here we presented only some axioms that seem to be sound with respect to the operation.

Thomas Meyer and collaborators (2002) propose a notion of contraction (systematic withdrawal) that is (according to Rott and Pagnucco 1999, Appendix B) also a particular case of a contraction operator obeying Inclusion. Meyer et al.'s contraction can be defined as follows: $[K \div A] = [K] \cup C \cup (R^x \cap [\neg A]$, with $\rho^+(\neg A) = x$, and $A \in K$, where $C = S^x - R^x$. The most interesting axioms proposed by Meyer et al. are:

($\div$7) If $C \in K \div (A \wedge C)$, then $C \in K \div (A \wedge C \wedge B)$

($\div$8) If $A \in K$, $A \vee B \in K \div A$, and $B \notin K \div A$, then $A \in K \div (A \wedge B)$

($\div$9) If $A \notin Cn(\emptyset)$, and $B \in K \div A$, then $A \notin K \div (A \wedge B)$

($\div$7) provides a very weak version of Antitony. ($\div$8) and ($\div$9) provide further constraints. ($\div$8) indicates conditions under which a formula should be retained, and ($\div$9) indicates further conditions under which a formula should be discarded.

These are just examples of some concrete contraction operations obeying Inclusion and the postulates of core contraction. But I would conjecture that a large class of so-called *withdrawal* operators could be classified parametrically as well as core operators obeying particular additional constraints. The core of the theory of information value is able to accommodate various well-known notions of contraction previously characterized without appealing to decision-theoretic tools.

Before embarking in an analysis of sequential change, I summarize some of the main points discussed in this section as well as in the previous section. We introduced first three core postulates (WM, EWM, and WIE) that articulate the notion of damped informational value used in contraction. These postulates propose a compromise between some of the structural properties of probability-based measures of informational value (encoded via WM and

EWM), on the one hand, and the recommendation of breaking ties by moving to a "skeptical" position of suspense (encoded via WIE), on the other hand. The logical commitments of these postulates are reflected (at least) by the postulates ($\div$1) to ($\div$6).

To what extent are these postulates *rationality constraints*, in the sense that changing one's view not in accordance with them would be tantamount to behaving irrationally? Unlike many other accounts of belief change, the core postulates are constraints on *value* functions. So, any person endorsing the goals and values that the core postulates intend to articulate should obey the corresponding postulates on pain of irrationality. But an agent whose values are restricted only by a general theory of utility, say, by the constraints imposed by the von Neumann-Morgenstern theory of expected utility, might nevertheless change his or her view not in accordance with these postulates without lapsing into irrationality (at least not under the point of view of the theory presented here). So, Levi's theory is certainly motivated by decision-theoretic considerations, but the theory diverges from other attempts to derive principles of theory change from preferences that are also based on decision theory but that do not assume the particular view of the goals of inquiry that Levi endorses (Morris 1997; Asheim and Søvik 2003; Collins 2002).

This being said, it is important to have a clear view of the real commitments of Levi's theory. As I explained above the postulate of Antitony is certainly not consensual among scholars working in this field. Some strongly endorse it, while others reject it with equal emphasis. But, as I painstakingly tried to explain above, doubts about Antitony should not be conflated with doubts about the core postulates of damped informational value. The endorsement of Antitony depends on more global considerations concerning the unification of the theory of theory change by a small set of mutually coherent maxims. For example, anyone endorsing the maxim that when A is given up from a theory K one should preserve all sentences better entrenched than A should endorse the full force of Min and therefore Antitony (as long as he or she also endorses the maxim that losses of information value should be kept minimal). Arguments backing this statement are given both in Levi (2004) and Arló-Costa and Levi (2004). But, of course, anyone having a different view of the role of entrenchment in theory change might manifest doubts about the extent to which Min holds, and this might lead to a weaker theory of contraction.

I think that one of the main contributions of Levi's theory of belief change is to make clear that changing one's view (deliberately) involves a choice and that this choice should be treated, as any other choice, with the tools of decision theory. If one sees things under this point of view, then it is clear that values and goals should enter into the analysis of change and that they should contribute to this analysis with importance equal to other purely doxastic aspects of cognition. Unfortunately, it is this very point that has not been taken seriously by researchers working intensely in this area for more than twenty-five years (philosophers, economists, computer scientists, etc). Since the mid-1970s there has been a considerable amount of work in this field (mostly done in computer science), but the dominant aspect of this work is exactly that changing one's view concerns only the way that purely doxastic representations are updated with *given* inputs. Many of these models have been strongly influenced by sophisticated accounts of how probability should be updated. To a large extent there are a variety of theories because there are a variety of doxastic representations. They could be sentential representations or propositional or probabilistic or graphic. They could be representations via finite sets of sentences, via theories or ordinal conditional functions, via probability (standard or lexicographic or infinitesimal), via Bayesian networks, and so on. But if, say, the representation is via theories, the theory of expansion is usually considered the most trivial part of belief change. One just adds the sentence to the theory and takes the logical closure. The question of "why expand with this particular input?" is never asked. It is usually presupposed that some inductive procedure has recommended doing so. In contrast the theory of expansion for Levi is far from being trivial. It presupposes a form of abduction (identifying a set of potential answers to a particular problem) and it also involves the application of the principles of rational choice to appropriate cognitive goals in order to choose among the potential answers identified by the abductive procedure.

By the same token, the question of: "what to remove?" in coerced contraction should be seen as a choice according to Levi, and the value of information plays a central role here as well, as we have seen, although it has to be articulated in a manner slightly different from the case of expansion (damped informational value). Researchers not fully endorsing this view sometimes talk about *preferences*, but it is not completely clear to what extent this talk is regimented by decision-theoretic considerations. Usually, it is not explicitly constrained by such considerations.[4]

Once value is recognized as a central component of the theories of expansion and contraction, we have the further problem of articulating more precisely the notion of informational value. The core postulates (or a suitable

[4] Hans Rott has connected the notion of preference with principles of rational choice by utilizing selection functions or choice functions of the sort that Amartya Sen extensively studied. See Levi (2004) for a critical approach to this material. See also Arló-Costa (2005).

extension of these postulates; see note 2) offer an answer to this question. And the theory enforcing Min offers a more comprehensive answer encompassing the role of entrenchment in contraction. But the decision-theoretical approach is quite rich representationally and it certainly admits for the parametrical accommodation of rival views. As a matter of fact, one of the main contributions of the AGM approach, aside from the peculiarities of the algebraic semantics it first offered, was to apply the axiomatic method to this area of inquiry. In fact, the proposal of axioms for belief change operators advanced the state of the art in this field beyond Levi's own contribution (Levi 1980). And some of the work done by Levi after the publication of AGM's theory focused on discussing the correctness of the AGM axioms from the point of view of his decision-theoretic account of belief change. This work has led to progress via the more precise statement of Levi's theory and to subsequent revisions of it. Levi's main goal is to determine the shape of the "official" theory of belief change flowing from the principles constraining the type of information value suitable to study contraction (damped information value). And some of the recent work done by Levi in collaboration (specifically, Arló-Costa and Levi 2004) focuses on finding the exact axiomatic base that characterizes his decision-theoretic proposal, as formulated in its revised form (Levi 2004). But there are obvious extensions of the theory that remain both conceptually and formally unexplored. The next section, for example, utilizes standard operators of information value in order to build an account of iterated supposition. This section has been devoted instead to a preliminary analysis of the notion of core contraction and to a preliminary classification of some notions of contraction already studied in the literature as various operators of core contraction. Under a decision-theoretical point of view the operator of mild contraction remains the axiomatization of the theory of information-value contraction with the best justification. Core contraction has, nevertheless, the appeal of offering the weakest theory of information value contraction (without imposing Min beyond what is strictly necessary in order to enforce WIE). One of the open questions concerning core contraction is what is the notion of entrenchment that can be used to represent it and to what extent this notion is philosophically adequate.

3. SEQUENTIAL CHANGE

The study of general principles regulating sequential change has been the most active area of research in the area of belief change during the last ten to fifteen years. The syntactic expression of the AGM theory permits the formulation of iterated axioms, and despite a widespread belief to the contrary, the standard formulation of the AGM theory does entail constraints on possible theories of iteration (see the arguments presented in Arló-Costa 1999). Nevertheless, there are rival presentations of the theory of belief change where iteration can be more easily expressed and studied. The paradigmatic example is the theory of *ordinal conditional functions* of Spohn or a slight variation of it under the name of *ranking theory*. The majority of the researchers who studied iteration in recent years have endorsed this view, with or without explicit mention of the pioneer work by Spohn. Let's stop here for a moment to consider the basis of Spohn's proposal (Spohn 1988, 2002). A *ranking function* κ is a function from $\mathcal{M}$ (see section 1) to the set of extended non-negative integers $\mathcal{N}^{+} = \mathcal{N} \cup \infty$, such that $\kappa(w) = 0$, for some $w \in \mathcal{M}$. For each proposition $P \subseteq \mathcal{M}$, the rank $\kappa(P)$ of P is defined by $\kappa(P) = \min\{\kappa(w) : w \in P\}$ and $\kappa(\emptyset) = \infty$. According to Spohn, ranks are best interpreted as *grades of disbelief.* $\kappa(P) = 0$ says that P is not disbelieved at all. It does not say that P is believed; this is rather expressed by $\kappa(P^c) > 0$, that is, that non-P is disbelieved (to some degree). The set $C(\kappa) = \{w\colon \kappa(w) = 0\}$ is called the *core* of κ and C_κ is the strongest proposition believed (to be true) in κ. So, if A^c is believed to be true in κ, one way of representing the contraction of A^c from $C(\kappa)$ is to take the union of $C(\kappa)$ with the set of least disbelieved A points, that is, $\{w\colon \kappa(w) = \kappa(A)\}$. This is a simple way of defining an AGM contraction in this setting.

κ can also be *revised* by A, for example, by implementing the following algorithm: $(\kappa^{\bullet}A)(w) = \kappa(w) - \kappa(A)$, if w is an A-point, and $\kappa(w) + 1$ otherwise. Notice that this algorithm gives us centrally a rule for revising entire *rankings*, and only an indirect rule for revising belief sets or theories (by tacitly revising ranking cores). So, if we are given a rule of this type, an initial ranking, and a sequence of inputs (which could contain mutually inconsistent inputs), we are guaranteed to arrive to a final ranking after processing the data. The core of the initial ranking suffers a series of transformations until the last transformation yields the core of the final outputed ranking, which is the strongest proposition believed after the sequence is processed. There are many transformation rules of the type proposed above and there is also a lively debate as to which one should be endorsed (if any). Symmetry arguments (Darwiche and Pearl 1997), computational arguments (Goldszmidt and Pearl 1992), and learning theoretic arguments (Kelly 1998) have been proposed, leading to different standards.

Much of this work is difficult to relate to Levi's work. Ranking systems and shell systems are vastly different from a conceptual point of view, and they are also different formally. I present nevertheless one of the recent proposals for iteration given that this is a proposal that does have some (indirect) connections with Levi's work. The proposal, presented by Freund and Lehmann (1994) consists syntactically in adding the following axiom to the AGM axioms for revision (see Alchourrón et al. 1985 for a list of these axioms):

$$(K^*9)\, K^*A = (K + A)^*A$$

It is not difficult to give a model of this theory in terms of a slight modification of ranking systems. Let $\kappa_{\emptyset}$ be the ranking obtained from κ by assigning the zero rank to the empty set and adding a unit to the rank of every world in κ. For all theories K (including $K_{\perp} = Cn(\perp)$), we have the same ordering induced by a ranking function κ that has to be kept fixed through sequential changes. Once this generalized notion of rank is fixed, characterize * as follows. If $[A] \cap [K] \neq \emptyset$, then define $[K^*A] = [K + A] = [K] \cap [A]$. If $[A] \cap [K] = \emptyset$, define $[K^*A] = \kappa_{\emptyset}(A) = \{w\text{: } w \mid = A,$ and $\kappa_{\emptyset}(w) \leq \kappa_{\emptyset}(z)$, for any $z \neq w, z \mid = A\}$. Of course, in this case we have $[K_{\perp}{}^*A] = C(\kappa_{\emptyset}{}^{\bullet}A) = [K^*A]$, for any arbitrary K. Therefore when A is incompatible with K, we have $[(K + A)^*A] = [K^*A]$ as (K^*9) requires. And when A is compatible with K, we have $[(K + A)^*A] = [K + A]$.

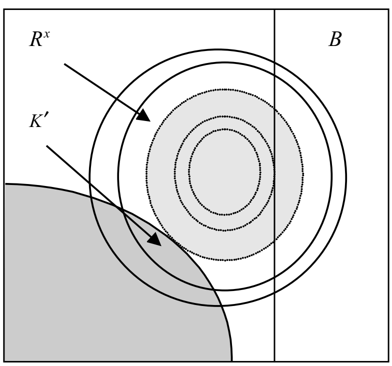
The main theorem of Freund and Lehmann (1994, theorem 2, sec. 4.1) establishes that there exists a bijection between the set of revisions that satisfy the postulates $K^*1 - K^*9$ and the set of consistency-preserving, rational relations, a form of nonmonotonic inference presented in Lehmann and Magidor (1988). The author's intuitive justification of this proposal goes as follows:

> When the agent receives new information ϕ, the agent revises its current belief K by either adding ϕ to its current belief set, if ϕ does not contradict K, or, if it contradicts K, *by forgetting about K altogether and adopting* ϕ *and all default assumptions that go with* ϕ *as its new belief set*. In the latter situation the agent adopts not only ϕ but all its usual, normal consequences also.

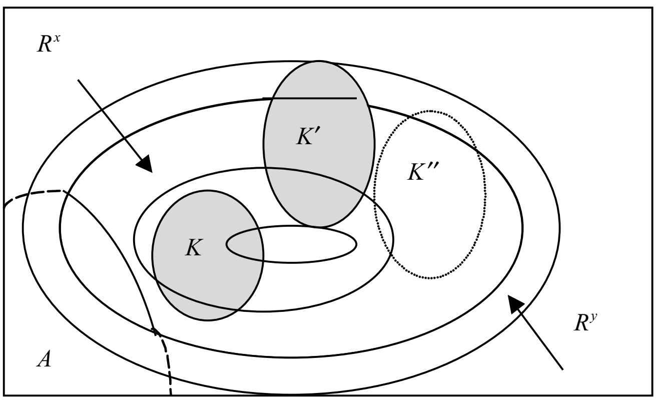
So this procedure characterizes a special type of revisions of belief sets performed under a fixed set of default assumptions, which are never revised. I argued (in Arló-Costa 1996) that there are weaker and more adequate belief revision theories mappable to rational relations (a point that Freund and Lehmann themselves noticed in the conclusion of their article). In spite of also having a divergent idea about the relationships between belief revision and nonmonotonic logic (see Levi 1996) Levi always manifested some sympathy for the view of iteration that flows from that of Freund and Lehmann, who attribute one of their axioms to Levi:

(L) If $\neg A \in K \cap K'$, then for any $C \in K \cap K'$, $C \in K^*A$ if and only if $C \in K'^*A$

The axiom is closely connected with Freund and Lehmann's theory. In fact, the authors prove that (L) and (K*9) are equivalent in the presence of (K*1) and (K*2), as long as the inconsistent set is allowed in the domain of the belief revision function. With some modifications the axiom will be useful for our purposes here. Notice that any decision-theoretical model of contraction requires having a value function V (relative to a theory K and a basic partition $\mathcal{B}$) constrained by certain principles. One possibility here is to study changes in belief determined while the value function initially given for K and $\mathcal{B}$ remains constant. A considerably more complicated option would be to study iterated change where the value function can be changed as well. This second option is more complex because it involves having a theory of value change. This is an interesting area of research almost completely unexplored so far. So, I restrict the present study to iterated changes of belief performed while the value function remains constant. Also I assume that *the underlying notion of contraction is the notion of mild contraction presented above*. Notice that even if the value function remains constant the associated system of shells need not remain constant through a sequence of changes (*because V* is constant). In fact, the ranks of the shell system are determined relative to a fixed theory of reference K. But say that the revision has been performed and that the current theory [K] changes to $[K'] = R^x \cap [A]$, where $\rho^-([K]) = x$. What is the shape of the system of shells for this new theory K'?



The figure can give the reader a more graphic idea of the situation. Let $[A]$ be depicted by the dark gray region. The intersection of $[A]$ with the rank R^x (indicated with an arrow) determines the revision K' of the theory K on which the shell system is based with A. Now of course the intersection of any set of partition cells $Y \subseteq R^x$ with K' continues to have value x. But as a matter of fact, the intersection of any set of cells $Z \subseteq S^x$ (including areas in the light gray region of the picture) with K' will continue to have value x as well. Min imposes this constraint. Suppose that now we revise K' with the sentence B, expressing the proposition $[B]$ indicated graphically in the picture by the rectangular region to the right of the picture. This could be done in two steps following the "Levi-identity." First contract $\neg B$ from K'. We can use a suitable modification of the initial shell system in order to perform the contraction. As the figure indicates, all the grading inside S^x becomes irrelevant. So, $[K' \div \neg B] = S^x$. Once the contraction is performed, then *expand* with B: $[K'^* B] = [K' \div \neg B] \cap [B] = S^x \cap [B] = [(K' \div \neg B) + B]$ where "+" denotes expansion. In general, if $\mathcal{S}$ is an initial system of shells for an initial theory K_0,[5] we can utilize $\mathcal{S}$ to define a system of shells for any other theory K, such that $[K] \subseteq 2^{\mathcal{D}}$. If $\rho^-([K]) = x$, then $\mathcal{S}_K = \mathcal{S} - \{R^y\colon y \geq x\} \cup S^x$. Of course, this covers the particular case where $K = K_0{}^*A$, for $[A] \subseteq 2^{\mathcal{D}}$. Then we can just define, for A incompatible with K, $[K \oplus A] = \{W \in \cup\mathcal{D}\colon w \in [A]$ and $\rho'^+(w) = \rho'^+([A])$, where ρ'^+ is the defined with respect to $\mathcal{S}_{\mathrm{K}}\}$. Of course, this covers the particular case where $K = K_0 \oplus A$, for $[A] \subseteq 2^{\mathcal{D}}$. There are some degrees of freedom in defining the revision of K with A with respect to a shell system S for K_0, but the salient options seem to classifiable as particular cases of the previous proposal.[6]



[5] Here it would be convenient to have an initial shell system where $[K_0]$ is the innermost shell of this system.

[6] One could argue perhaps that rather than calculating revisions with respect to $\mathcal{S}_K$ one should first revise $\mathcal{S}$ with K in a different way (by taking into account the upper K-ranks, rather than the lower ranks) and then revise the output of this revision with A. If $\rho^+([K]) = x$, then $\mathcal{S}\bullet K = \mathcal{S} - \{R^y\colon y \geq x\} \cup S^x$. The new shell system is now a shell system for the theory K' corresponding to $R^x \cap [A]$. So we can define for A incompatible with K': $[K'^* KA] = \{w \in \cup\mathcal{D}\colon w \in [A]$ and $\rho'^+(w) = \rho'^+([A])$, where ρ'^+ is the defined with respect to $\mathcal{S}\bullet A\}$. The arguments presented above also apply for this kind of characterization of iteration.

So far we have considered revisions of theories with propositions incompatible with them. Otherwise, if the current shell system is $\mathcal{S}_K$ and $[A]$ is compatible with $[K]$, then $\mathcal{S}_K$ is mapped to $\mathcal{S}_{K \cap A}$, and $K + A$ is the new theory of reference.

The axiom L seems to fail in some situations under the previous interpretation of the iterated revision operator. Consider the situation depicted in the picture above. $[C] = [K \cap K'] \cup (R^x \cap [A])$. In this case it is clear that $[K \oplus A] = R^x \cap [A] \subseteq [C]$. But $[K' \oplus A] = S^y \cap [A] \not\subseteq [C]$, given that the greatest rank for $\mathcal{S}_{K'}$ is $y < x$. Here it is a modification of the axiom that seems to circumvent counterexamples for the operator $\oplus$:

(L') If $\neg A \in K \cap K'$, then for any $C \in K \cap K'$, $C \in K \oplus A$ if and only if $C \in K' \oplus A$, as long as $C \in (K \cap K') \oplus A$.

More positively, there is also an axiom that seems to capture most of the salient features of iteration in this setting:

(**Fixity**) If $\neg A \in K$, A, $B \in \boldsymbol{L}$ and $\neg A \in (K \div \neg A) + B \neq \boldsymbol{L}$, then $((K \div \neg A) + B) \oplus A = K \oplus A$.

In the previous diagram, K'' is an example of a possible theory $((K \div \neg A) + B)$, with $[B] = [((K \div \neg A) + B)]$. Freund and Lehmann's proposal is radical in discounting the role of the background theory in revision. In fact, for any two theories K, K', as long as $\neg A \in K$ and $\neg A \in K'$, we have that the revisions of K and K' with A coincide. This situation is not verified in the account of iteration we are studying (when V is kept fixed), but as Fixity indicates, as long as A is rejected both in K and K'' and both $[K]$ and $[K'']$ are nonempty propositions included in $[K \div \neg A] \cap [\neg A]$, then $K \oplus A = K'' \oplus A$. Moreover, in these conditions we also have $K \div \neg A = K'' \div \neg A$. With more generality, the fact that A is rejected both in K and K'' guarantees that either $K \oplus A \subseteq K'' \oplus A$ or $K'' \oplus A \subseteq K \oplus A$. The coincidence of the A-revisions of any two theories K and K' as long as A is rejected in both is in Freund and Lehmann's account a consequence of the fact that there is only one default ordering for all theories. Such a radical result does not hold in the theory we are considering here, but we still have one fixed value function for all theories. And this leads to some of the coincidence results we listed.

There is an important issue related to the interpretation of iterated change that has not been considered yet. The iterated changes in question can be interpreted either synchronically or as bona fide rules for diachronic change. Levi's view on this issue, manifested in his rejection of "dynamic" Dutch Books, is well known. So, we shall assume here that the iterated changes we are studying are pre-compiled changes of view calculated "ex ante," rather than rules for temporal change, in such a way that an agent might fail to implement a precompiled change without lapsing into irrationality. It should be noted in passing that supposing that I come to believe A (and then B, etc., as various pieces of data in a given data sequence) might not be considered equivalent to supposing A for the sake of the argument (and then B, etc.). But even if the notion of iterated change were understood only in the latter sense, our counterexample has force. That this is so might not be immediate because Levi has rejected the "Levi-identity" in his analysis of suppositions for the sake of the argument. Supposing that A is the case from the agent's point of view requires, according to Levi, opening the agent's mind both with respect to $\neg A$ and to A and then expand with A. So, the revision according to Ramsey (as Levi calls it (1996)) is defined as follows: $K®A = ((K \div \neg A) \div A) + A$, where $K©A = (K \div \neg A) \div A$, that is, contraction according to Ramsey. Now notice that all the examples presented graphically in the previous picture apply also to Ramsey revision. The interest of Ramsey revision for the analysis of conditionals is determined by the fact that, for example, $K®\neg A \neq K$ even when $\neg A \in K$. But in our example, $K®A = K \oplus A$, and the same applies to K'' and to $K \cap \mathrm{K}'$. So adopting the unmodified axiom L does not seem to be appropriate to axiomatize the kind of iterated suppositions involved in Ramsey revision either. And Fixity seems to circumvent this problem for Ramsey revision as well. Notice that Fixity is a hybrid axiom combining contraction and revision operators. But, of course, the axiom can be purely expressed in terms of $\oplus$(®) by adopting the so-called Harper identity if revision can be defined in terms of contraction using the Levi identity. If, instead, we utilize Ramsey revision and Ramsey contraction, I would propose to define Ramsey contraction in terms of Ramsey revision via the following identity:

$$K©A = K®\neg A \cap K®A$$

One of the cases where there is divergence between recent proposals for iteration concerns the case where an agent receives (sequentially) contradictory reports about one and the same situation. Cases of this sort include situations where an agent receives information through a noisy channel or from a reliable but not infallible oracle. Some accounts have proposed axioms according to which when two contradictory pieces of evidence arrive, the last one prevails; that is, the second piece of evidence alone would yield the same state of belief. The axiom is known as C2 according to the terminology used in Darwiche and Pearl (1997). This axiom is not obeyed by $\oplus$ in the present proposal. We have instead: (C) If $\neg A \in K$, then $[(K \oplus A) \oplus \neg A] = [K \div \neg A] \cap [\neg A]$. Only when $A \in K$, we of course have that $(K \oplus A) \oplus \neg A = K \oplus \neg$A. C2 is a controversial axiom, and its violation can hardly be counted as a problem for an account of iteration (see Arló-Costa and Parikh 1995). The other axioms proposed in Darwiche and Pearl (1997) are satisfied as well:

(C1) If A entails B, then $(K \oplus B) \oplus A = K \oplus A$

(C3) If $B \in K \oplus A$, then $B \in (K \oplus B) \oplus A$

(C4) If $\neg B \notin K \oplus A$, then $\neg B \notin (K \oplus B) \oplus A$

These axioms have to be weakened for Ramsey revision (e.g., C1 holds only for *A*, *B* that are either both rejected or kept in suspense relative to *K*). It is clear that the view of iteration that we are delineating here is different from many of the rival accounts studied in the literature. Even when Fixity is still true in Freund et al.'s theory, this account goes beyond what we deem permissible here in ignoring the role of the background theory for defining *K**A*, when *A* is rejected in *K*. On the other hand, the accounts based on Spohn's ideas utilize ranking systems, which are structurally different from shell systems. And Fixity is not satisfied in most variants of this view. In fact, this view seems to be based on a completely different idea according to which the "preferences" or the plausibility ranking of an agent changes every time that an input is processed. Recent accounts of iteration in computer science (Delgrande 2004) have criticized the Spohnian view for what seems an at-least-sometimes inappropriate notion of locality in implementing iterated revisions (according to which the most recently discarded information is always most plausible with respect to the current knowledge base). And this has motivated researchers to propose the construction of an iterated theory of change on the basis of assigning a fixed distance between possible worlds and requiring that the set of worlds together with this distance form a metric space (Lehmann, Magidor, and Schlechta 2001; Delgrande 2004). Again, this kind of construction is quite different from the one presented here even when most of the motivating examples utilized by Delgrande or Lehmann et al. can be dealt with in the framework we are considering here.[7]

[7] In a typical counterexample an agent believes something of the form $A \wedge B$ and he or she is notified next that $\neg A \wedge \neg B$ (a situation that he or she considers most plausible among $\neg(A \wedge B)$ options). In approaches such as that of Darwiche and Pearl (1997), if the agent is then given the information that *A*, he or she will lose the information that $\neg B$. Delgrande, for example, argues that in certain situations this is undesirable. He considers the case where *A* and *B* are irrelevant (*B* stands for "Sherlock Holmes existed" and *A* for "The temperature is 18F"). But it is not clear what should be preserved or not when *B* stands instead for "it is snowing" – suspense about *B* seems advisable. It seems useful here not to confound problems by appealing to relevance considerations, which in Levi's account can be handled

Unlike some of the recently proposed theories of iteration, Fixity is derived in this account from the plausible assumption that the agent's *cognitive values* relevant in doxastic choice are persistent throughout a sequence of epistemic changes.

Levi abandoned (in Levi 1996) the Stoic view that the only legitimate conditionals are nonnested conditionals, and he tried to build the theory of iterated supposition needed in order to study iterated conditionals. He proposed there the notion of Ramsey revision presented above, although the underlying notion of contraction he used was not the theory of contraction discussed here (mild contraction). This notion was proposed after the publication of Levi's book. *For the Sake of Argument* also contains some initial ideas about iteration. The theory presented in this section intends to continue this effort by adopting the right underlying notion of contraction and by proposing a theory of iteration for it and for the associated notion of Ramsey revision. Of course, the assumption that the value function remains fixed throughout a sequence of changes is even more plausible for this suppositional interpretation of revision (and contraction).

REFERENCES

Alchourrón, Carlos, Gärdenfors, Peter, and Makinson, David. 1985. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions." *Journal of Symbolic Logic* 50: 510–30.

Arló-Costa, Horacio. 1996. "Belief Change and Suppositional Reasoning: Knowledge Representation, Defeasibility and Conditionals." Ph.D. thesis, Columbia University.

separately via the selection of adequate partitions. Counterexamples in terms of relevance can be presented to all notions of belief change, including the one we are articulating here, to Spohn's, and so on. This is why it is better to treat the issue of relevance separately (as in Parikh 1999) or via partitions.

But, more to the point, it is easy to check that in the extension to iteration we are considering here it is possible to preserve $\neg B$ after processing A in the second change if considerations in terms of information value require to do so. It is also possible to be in suspense about B, and finally, it is also possible to revert to B after the second change. All depends on the shape of the underlying value function. It might be instructive to consider the case where $\neg B$ is preserved after processing A in the second change. Assume that the current shell is $\mathcal{S}_K$ where $K = Cn(A \wedge B)$, and where the innermost rank r of the initial shell $\mathcal{S}$ is occupied by $K' = Cn(\neg A \wedge \neg B)$, rank $r - 1$ contains only $Cn(A \wedge \neg B)$, and finally rank $r - 2$ is $Cn(A \wedge B)$ – in such a way that $\mathcal{S}_K$ has maximal rank $r - 2$. $(\mathcal{S}_K)_{K'}$ reverts to $\mathcal{S}$. After a new change with A, the new theory of reference is $K'' = Cn(A \wedge \neg B)$, and the new shell is $\mathcal{S}_A$ (with maximal rank $r - 1$). If the complaint against the Spohnian account implemented in Darwiche and Pearl (1997) is that it rules out "a priori" some iterated changes that one might nevertheless consider permissible, then the point seems to be well taken. But the extension to iteration presented here is able to circumvent these problems without appealing to the notion of (mereological) distance between possible worlds, a notion that seems difficult to motivate either epistemologically or decision-theoretically.

Arló-Costa, Horacio. 1999. "Belief Revision Conditionals: Basic Iterated Systems." *Annals of Pure and Applied Logic* 96: 3–28.

Arló-Costa, Horacio. 2005. "Rationality and Value: The Epistemological Role of Indeterminate and Agent-Dependent Values." In Vincent F. Hendricks (ed.), "8 Bridges between Mainstream and Formal Epistemology." Special issue of *Philosophical Studies*.

Arló-Costa, Horacio, and Isaac Levi. 2004. "A Complete Characterization of a Notion of Contraction Based on Information-Value." *Proceedings of the 10th International Workshop on Non-Monotonic Reasoning*, pp. 25–34. http://www.pims.math.ca/science/2004/NMR/papers.html.

Arló-Costa, Horacio, and Rohit Parikh. 1995. "On the Adequacy of C2." Research Note, Columbia University and CUNY-GC.

Asheim, Geir B., and Ylva Søvik. 2003. "The Semantics of Preference-based Belief Operators." Working Paper, University of Oslo, Economics, abstract in TARK.

Booth, Richard, Samir Chopra, Aditya Ghose, and Thomas Meyer. 2003. "Belief Liberation (and Retraction)." In Moshe Tennenholtz (ed.), *Proceedings of the Ninth Conference TARK 2003: Theoretical Aspects of Rationality and Knowledge*, pp. 159–72.

Collins, John. 2002. "Belief Revision Derived from Preference." Talk presented at meeting in Philosophical Logic and Applications, CUNY Graduate Center, September 27.

Darwiche, Adnan, and Judea Pearl. 1997. "On the Logic of Iterated Belief Revision." *Artificial Intelligence* 89, nos. 1–2: 1–29.

Delgrande, James. 2004. "Preliminary Considerations on the Modeling of Belief Change Operators by Metric Spaces." *Proceedings of the 10th International Workshop on Non-Monotonic Reasoning*, pp. 118–26. http://www.pims.math.ca/science/2004/NMR/papers.html.

Freund, Michael, and Daniel Lehmann. 1994. "Belief Revision and Rational Inference." Technical Report TR 94-16, Hebrew University, Computer Science.

Goldszmidt, Moises, and Judea Pearl. 1992. "Rank-based Systems: A Simple Approach to Belief Revision, Belief Update and Reasoning about Evidence and Actions." *Artificial Intelligence* 84, nos. 1–2: 57–112.

Hansson, S. O. 1999. *A Textbook of Belief Dynamics: Theory Change and Database Updating*. Dordrecht: Kluwer Academic.

Hansson, S. O., and E. J. Olsson. 1995. "Levi Contractions and AGM Contractions: A Comparison." *Notre Dame Journal of Formal Logic* 36: 103–19.

Kelly, Kevin. 1998. "Iterated Belief Revision, Reliability, and Inductive Amnesia." *Erkenntnis* 50: 11–58.

Lehmann, D., and Magidor, M. 1988. "Rational Logics and Their Models: A Study in Cumulative Logics." Department of Computer Science, Hebrew University, Jerusalem, Israel. Technical report TR-88-16.

Lehmann, D., M. Magidor, and K. Schlechta. 2001. "Distance Semantics for Belief Revision." *Journal of Symbolic Logic* 66, no. 1: 295–317.

Levi, Isaac. 1980. *The Enterprise of Knowledge*. Cambridge, Mass.: MIT Press.

Levi, Isaac. 1991. *The Fixation of Belief and Its Undoing*. Cambridge: Cambridge University Press.

Levi, Isaac. 1996. *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference, and Nonmonotonic Reasoning*. Cambridge: Cambridge University Press.

Levi, Isaac. 2004. *Mild Contraction: Evaluating Loss of Information Due to Loss of Belief*. Cambridge: Cambridge University Press.

Meyer, Thomas, Johannes Heidema, Willem Labuschagne, and Louise Leenen. 2002. “Systematic Withdrawal.” *Journal of Philosophical Logic* 31, no. 5: 415–43.

Morris, S. 1997. “Alternative Notions of Belief.” In M. O. L. Bacharach, L.-A. Gerard-Varet, P. Mongin, and H. S. Shin (eds.), *Epistemic Logic and the Theory of Games and Decisions*, pp. 217–33. Dordrecht: Kluwer.

Parikh, Rohit. 1999. “Belief Revision and Language Splitting.” In *Proceedings Logic, Language and Computation*, Lawrence S. Moss, Jonathan Ginzburg, and Maarten de Rijke (eds.), pp. 266–78. CSLI Stanford, Calif.: Center for the Study of Language and Information.

Rott, Hans. 1991. “Two Methods of Constructing Contractions and Revisions of Knowledge Systems.” *Journal of Philosophical Logic* 20: 149–73.

Rott, Hans. 2000. “Two Dogmas of Belief Revision.” *Journal of Philosophy* 97: 503–22.

Rott, Hans, and Maurice Pagnucco. 1999. “Severe Withdrawal (and Recovery).” *Journal of Philosophical Logic* 28: 501–47.

Spohn, Wolfgang. 1988. “Ordinal Conditional Functions: A Dynamic Theory of Epistemic States.” In W. Harper and B. Skyrms (eds.), *Causation in Decision, Belief Change and Statistics*, vol. 2, pp. 105–34. Dordrecht: Kluwer.

Spohn, Wolfgang. 2002. “Laws, ceteris paribus Conditions, and the Dynamics of Belief.” *Forschungsberichte der DFGForschergruppe Logik in der Philosophie* 84.

15

Deciding What You Know

Mark Kaplan

I

You are driving with your partner late in the afternoon of a hard day. She asks you, "Do you know if the bank is open tomorrow? It's a Saturday." "Yes, it will be open," you reply. "Thing is," she continues, "we're both tired. It would be nice to quit for today and go to the bank tomorrow. Mind you, if you're wrong and the bank isn't actually open tomorrow, we won't be able to make the mortgage payment on the house and we will lose it to our lender." Would it be entirely surprising if you replied, "I take it back; I don't know if it will be open – let's go now"?[1]

I think we can all agree that it would not be entirely surprising. And I think we can also all agree that, all the same, the reply leaves us a bit uncomfortable. The thought is this. Whether you know the bank is open on Saturday depends entirely on how you are epistemically situated with respect to the proposition that it will be open Saturday. What the stakes are, should you act as if the bank will be open Saturday, affects not a bit your epistemic situation with respect to that proposition. So, your epistemic situation with respect to the proposition that the bank will be open Saturday cannot properly be thought to change when you learn of the stakes that ride on the bank's being open then. And so, the thought is, your way of determining whether you know that the bank will be open Saturday, sensitive as that way is to your change in opinion as to what those stakes are, is (in at least one of its deployments) quite defective.

I would like to thank Jonathan Weinberg and Joan Weiner for comments on this chapter; Matthew McGrath for helpful discussion of a paper I read at Texas A&M that contained some of the ideas I develop here; and Adam Leite for comments on the chapter and extensive conversation about the doctrines it advances.

[1] This example is a variation on one featured in Keith DeRose, "Contextualism and Knowledge Attribution," *Philosophy and Phenomenological Research* 52 (1992): 913–29, 913.

This, it seems to me, is an attractive line of thought – that whether a person knows that *P* should, in general, be unaffected by what is at stake if he or she acts on *P* in his or her current circumstances. It is attractive to think that, to tell whether you know, you need pay attention only to epistemic matters. No doubt this explains the fact that the huge literature on the analysis of knowledge – a literature devoted to describing the conditions that determine whether one knows that *P* – has paid little or no attention to how the stakes in acting on *P* might figure in the decision. The thought has been that they don't.

But, attractive though it may be, I am convinced this line of thought is mistaken. What I would like to do is to explain why – and to offer another, more adequate way to think about the bank case and about what it teaches us about knowledge.

II

Why do we find unsurprising your imagined reaction to the news that, if you don't go to the bank today and it is closed on Saturday, you lose your house? The reason, I maintain, is that it is because there is the following fact about knowledge: You may not think of yourself (i.e., you open your position to criticism[2] if you think of yourself) as knowing that *P* if, at the same time, you are unwilling to act as if *P* is true.[3] That is, you may not think of yourself as knowing that *P* if, at the same time, you acknowledge that, among a number of options open to you, there is one that will have a consequence if *P* that no other option can better, yet you are unwilling to take that option. (In this respect, knowledge is unlike belief. There is nothing the least bit objectionable about your believing the bank is open Saturday but being unwilling to behave as if it is.[4])

This fact is crucial to the explanation of why, on learning of the dire consequences of acting as if the bank will be open and being mistaken, you would withdraw your claim to know. You withdraw your claim to know that

[2] There is a distinction between opening your position to criticism by doing a particular thing and opening *yourself* to criticism by doing that thing. See my "Epistemology Denatured," in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein (eds.), *Midwest Studies in Philosophy XIX: Philosophical Naturalism* (Notre Dame: Notre Dame University Press, 1994), pp. 350–65, esp. sec. V.

[3] Much the same point is made by Jeremy Fantl and Matthew McGrath in "Evidence, Pragmatics and Justification," *Philosophical Review* 111 (2002): 67–94, 72, a paper that, like this chapter, explores the consequences of taking this point seriously.

[4] Contrast ibid., p. 77.

the bank will be open Saturday (at least in part) because you find yourself unwilling to act as if it will be open in the light of the dire consequences that you have just learned will ensue should it actually be closed. Whether or not you are *right* to withdraw that claim to know (or were right to make a claim to know in the first place), it is the fact that you may not regard yourself as knowing the bank will be open if you are unwilling to act as if it will be open that explains why you withdraw your claim to know on learning what you do. Absent this fact about knowledge, why would you feel any pressure whatsoever – why would we be so unsurprised when it turns out that you want – to withdraw your claim to know?

But, if this is right, then the attractive thought rehearsed above – the thought that, to determine whether you know that *P*, you need not in general concern yourself with what is at stake if you act as if *P* – leads immediately to the following result: You may regard yourself as knowing that *P* (you open your position to criticism unless you regard yourself as knowing *P*) only when you are willing to bet all and everything, for even the most modest gain, on *P*'s truth. For suppose you regard yourself as knowing that *P* when it is not the case that you are willing to bet all and everything on *P*'s truth. Then there is a way for the consequences of acting as if *P* when *P* is false to be rendered dire enough that you will then be unwilling to act as if *P*. In this event, given that you may regard yourself as knowing that *P* only if you are willing to act as if *P*, you may no longer (you open your position to criticism if you continue to) regard yourself as knowing that *P*. That is to say, you must behave precisely as we imagined you behave in the bank case. On pain of opening your position to criticism, you must change your decision as to what you know despite the absence of any epistemic change in your circumstances. You must allow the decision as to what you know turn on the fact that the stakes have been raised.

The trouble with this result – the trouble with the result that you may regard yourself as knowing *P* only if you are willing to bet all and everything on *P*'s truth – is that it seems to lead inescapably to the conclusion that we know far less than we thought. Timothy Williamson is surely not alone in wanting to say, "I know plenty of things without my being willing to bet my house on them."[5]

No one has worked more valiantly, and more ingeniously, than Isaac Levi to blunt the blow inflicted by this result. In a series of books and articles

[5] Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000), p. 86. For all that, Williamson there ends up (unintentionally) committed to a view of knowledge that is at odds with the sentiment expressed in the quoted statement – as I argue in "Who Cares What You Know?" *Philosophical Quarterly* 53 (2003): 105–16.

dating back to the early 1960s, Levi has labored to provide an account of how to decide what you know that will (1) respect the purely epistemic character of the considerations that determine whether a proposition deserves a place in your corpus of knowledge,[6] (2) accept the result that you may not regard yourself as knowing that *P* without being willing to bet all and everything on *P*'s truth, yet (3) provide a rationale for allowing your corpus of knowledge to be sizable.

How has Levi tried to do all this? Not by in any way soft-pedaling the commitment involved in your regarding yourself as knowing that *P*. For Levi, this involves not only assigning *P* a degree of confidence equal to 1 (hence a willingness to bet all and everything on *P*), but also a commitment not even to take into account, in decision making, of what the consequences are if not-*P*.[7] Rather, Levi has sought to distinguish this pair of commitments from a number of others with which it might be conflated, and for whose shortcomings the commitments Levi wants to associate with knowledge might otherwise be blamed.

In this, Levi has been remarkably resourceful. He has pointed out that the doctrine linking knowing that *P* with a willingness to bet all and everything on *P* is not in any way committed to incorrigibilism. It is one thing to be willing to bet all and everything on *P*; it is another to be unwilling, should the circumstances be right, to change your mind. It has, to be fair, turned out to be no mean feat to provide a convincing account of how such a change of mind might be rationalized.[8] But the fact remains: The doctrine that links knowing that *P* with a willingness to bet all and everything on *P* is not, *eo ipso*, committed to incorrigibilism. And it cannot be called to account for being incorrigibilist.

He has also pointed out that there is nothing in the doctrine that requires you to admit into your corpus of knowledge only propositions that are necessitated by your evidence. True enough, Levi notes, if all that mattered to the question of what to admit were the aim of avoiding of error, no evidence that failed to necessitate *P* would be sufficient to warrant *P*'s inclusion in that corpus. For, if error avoidance were all that mattered, then to include *P* when your

[6] Levi's worries about denying the autonomy of inquiry's aims date to "Must the Scientist Make Value Judgments?" *Journal of Philosophy* 57 (1960): 345–57. In *Gambling with Truth* (New York: Alfred A. Knopf, 1967), he champions their autonomy. See pp. 16–19.

[7] Note that one might have a degree of confidence in *P* equal to 1 without having this second commitment. Suppose *P* is "Ticket 1 will lose" where this ticket is one ticket in a countably infinite fair lottery.

[8] I discuss some of the difficulties in a review in *Philosophical Review* 92 (1983): 310–16, of Levi's *The Enterprise of Knowledge* (Cambridge, Mass.: MIT Press, 1983).

evidence does not necessitate *P* would be to run (what you yourself regard as) a real risk of importing error into your corpus for no gain.

But, Levi argues, error avoidance is not all that matters. That your corpus be contentful, that it offer explanatory power – these are also desiderata when deciding what to include in your corpus. And they are desiderata that, in the normal course of events, you simply cannot secure except at the cost of running the risk of importing error into your corpus. So it is that evidence that does not necessitate *P* can nonetheless warrant the import of *P* into your corpus of knowledge: The option of importing *P* need only offer, among the options of importing *P*, importing not-*P*, and importing neither *P* nor not-*P*, the best mix of error avoidance and content acquisition. That is to say that, insofar as the risk of error in, and the content delivered by, importing *P* admit of measurement – and so their relative epistemic utility admits of measure – evidence that does not necessitate *P* will warrant the import of *P* into your corpus of knowledge provided the import of *P* bears greater expected epistemic utility than the other two options: importing not-*P* and importing neither.

And yet Levi's effort does not succeed. There is still the stubborn fact that we want to join Williamson in thinking that there are plenty of things we know that we wouldn't be prepared to bet our house (our lives) on. Indeed, things are worse than that, given Levi's view. Levi – rightly, in my view – sees degrees of confidence as being open to criticism unless they satisfy the axioms of the probability calculus. That calculus requires that the conjunction of any set of propositions with probability 1 itself receives probability 1. The upshot of Levi's view, then, is that you must harbor (your position is open to criticism unless you harbor) a degree of confidence equal to 1 in the conjunction of all the propositions you regard yourself as knowing: You must be willing to bet all and everything that everything you know is true. It is hard to see how the willingness to make such a bet can be warranted unless the standards for admitting propositions into your corpus of knowledge are very much higher than those we ordinarily employ – that is, unless we count ourselves as knowing very much less than we ordinarily do.

To be sure, Levi has, in his account of the aims of inquiry, a story as to how the willingness to make such a bet can be warranted even for a corpus of knowledge as sizable as the ones with which we ordinarily credit ourselves: Our search for content and explanatory power provides a rationale for introducing propositions into our corpora even when these propositions are not necessitated by (or even maximally probable given) our evidence. But it is one thing to maintain that there must be a place in inquiry in which only epistemic considerations matter. It is another to maintain that, in our lives, epistemic desiderata take precedence over all others. The former is, as I have admitted, an attractive thing to maintain. The latter is not. Yet it is the latter to which Levi's account is committed. He would have us decide to admit propositions into our corpora – he would have us undertake a willingness to bet all and everything on their truth – solely on the strength of how well their admission serves the aims of error avoidance and (explanatory) power. The fact that admitting *P* will gain us power is to weigh with us; the fact that we stand to lose our house (our lives) if we undertake the willingness to bet all and everything on *P* that comes with the admission of *P* is to weigh not at all. But this would seem to elevate relief from agnosticism (Levi's early, wonderful expression for what a contentful corpus provides an inquirer[9]) to the status of a good that, apart from the good of error avoidance, is lexicographically prior to all other goods. And this is hard to swallow. While it would be truly nice if there were lots of propositions we were willing to bet all and everything on, it is nothing to risk our lives for.

This undue emphasis on the value of relief from agnosticism has an epistemic cost as well. Suppose you possess evidence that does not maximally probabilify *P* – that is to say, suppose that the degree of confidence you are warranted in assigning to *P* given the evidence is less than 1. Levi holds that, under the right circumstances, the desideratum that you harbor a contentful corpus of knowledge will warrant your admitting *P* into that corpus, that is, will warrant your assigning *P* a degree of confidence equal to 1. But this would seem to violate the Requirement of Total Evidence. By having you admit *P* to your corpus of knowledge, Levi would have you henceforth regard *P* as more probable (pretend that *P* is more probable) than it in fact was, given your evidence. He would have you deliberately lose critical information about your actual epistemic condition. It is the rough equivalent to a policy of throwing away your lab notes once you have decided what your experiment shows.[10]

As far as I am concerned, the conclusion is irresistible: The attractive line of thought I rehearsed in the initial section of this chapter must be rejected.

[9] See *Gambling with Truth*, p. 58.

[10] In this I express a fundamental disagreement with Levi, insofar as he holds (*Enterprise of Knowledge*, pp. 1–2) that

> [o]nce *X* has concluded that adding *h* to his corpus is justified as an improvement to his corpus and has implemented his decision, *h* ceases, for *X*, to be a hypothesis. It has become a premise, evidence, settled assumption, or part of the "background knowledge" to be used in subsequent inquiries into the credentials of other statements, as well as in practical deliberations aimed at moral, political, economic, or other practical objectives. Whether *h* is a theory, law, statistical claim, or observation report, and regardless of the grounds on which it has been added, its status as an item in *X*'s corpus has been settled *and the grounds on which it has been added no longer matter*. (emphasis added)

According to that line of thought, all you need to attend to, in order to tell whether you know that *P*, are purely epistemic matters about your circumstances. In particular, it is of no moment what the stakes are should you act as if *P* and should *P* turn out to be false. This line of thought must be rejected so long as we hold it to be a fact about knowledge that you may not regard (your position is open to criticism if you regard) yourself as knowing that *P* while, at the same time, you are unwilling to act as if *P*. (And as I argued earlier, it is hard to explain why you would behave as we imagine you do in the bank case – it is hard to explain why we are not surprised that you should behave as we imagine you do in the bank case – if this were *not* a fact about knowledge.) For, together with that fact, the attractive line of thought leads us to a most unattractive result: You may not regard yourself as knowing that *P* (you open your position to criticism so long as you regard yourself as knowing that *P*) unless you are also willing to bet all and everything on *P*'s truth. And, Levi's efforts notwithstanding, this result would force us to conclude that we know far less – I would argue drastically less – than we are accustomed to think.

What I would like to do now is sketch a way of thinking about how to decide what you know that abandons the attractive line of thought – and thereby allows us to see how reason can be at work in the behavior portrayed in the bank case. It is, to be sure, a way of thinking about how to determine what you know that makes that determination a great deal less important than it is, say, on Levi's account. On Levi's account, our decisions as to what we know determine relative to what we calculate how much confidence to invest in the propositions we do not know. But, as I have been arguing in this section, if we are to think of ourselves as knowing anywhere near as much as we think we do, we cannot possibly construe our decisions as to what we know as playing so fundamental a role in our lives. My burden will be to say something about what role they play instead. My conceit will be that, once we think about this matter properly, we can see that there is still truth in the attractive line of thought: That is to say, there *is* a place in the enterprise of inquiry in which only epistemic considerations matter. The error of the attractive line of thought is simply that it misidentifies that place.

III

It seems to me that the only way to think correctly about how we decide what we know, is by trying to capture our ordinary practice – that is, to try to say what constrains what we say and think and do, and think ourselves *right* to be saying and thinking and doing, in ordinary life (where, if part of our ordinary lives is spent pursuing the special sciences, then what we do there counts too).[11] It is precisely by that methodological standard that the attractive line of thought – and Levi's defense of the attractive line of thought – has come up short. I have already made a great deal of one of the constraints on thinking about what it is we know: You may not regard yourself (you open your position to criticism if you regard yourself) as knowing that *P* if, at the same time, you are unwilling to act as if *P*. And I have also made a great deal of there being a particular constraint that does *not* operate on thinking that you know. I have been arguing that it is not impermissible to let the stakes involved, should you act as if *P* and should *P* turn out to be false, affect your decision as to whether *P* belongs in your corpus of knowledge.

Taken together, these two make possible an easy rationale for your behavior in the bank case. Whether you can properly think of yourself as knowing that *P* depends, in part, on what decision problems are on your plate and what is at stake should you act as if *P* with respect to those problems and should *P* turn out to be false. In the bank case, you revise your view as to what the stakes are in your decision problem (whether to go to the bank today or on Saturday) and you come to appreciate how dire the consequences are of acting as if the bank will be open Saturday if, in fact, it is not. In so doing, you come properly to see that your grounds for thinking you know that *P*, good enough to warrant your being willing to act as if *P* in the circumstances as you previously thought them to be, are not good enough to warrant your acting as if *P* in the circumstances as you now appreciate them to be. So you withdraw your claim to know.

It is still an account, of course, on which you have made a mistake: You initially mistook what the true stakes were. But you have made no mistake in how you went about determining what you know. When it comes to determining what you know, the stakes enter the decision. Raise the stakes (or, as in the case at hand, learn that the stakes are higher than you thought), and some of what you claimed to know you will – as you should – cease claiming to know.[12]

[11] I defend this methodological commitment in "To What Must an Epistemology Be True?" *Philosophy and Phenomenological Research* 61 (2000): 279–304.

[12] But, if you must withdraw your claim to knowledge even in the case in which you have not mistaken what the stakes are – in the case in which the stakes have actually been raised – how are you in that case to regard your former claim to knowledge? Should you think of yourself as having known, but no longer? Should you think of yourself as never having known? I am inclined to think the latter. I am inclined to think that learning that the stakes have been raised is, in this respect, like learning something that undercuts a crucial bit of your evidence that *P*. It makes you think that you don't know now and didn't know before. (For a different verdict, see DeRose, "Contextualism and Knowledge Attribution.") But it is beyond the scope of the current sketch to settle the matter. It is also beyond the

This way of thinking about how to decide what you know also makes possible an easy rationale for the Williamson sentiment. Spoken or thought in a circumstance in which you are not faced with the decision as to whether to bet your house on *P*, it makes perfectly good sense to say that you know that *P*, but are not willing to bet your house on *P*. Your house is not, in fact, at stake in acting as if *P* in the circumstance. So your unwillingness so to act if your house is at stake in no way constrains your decision, in that circumstance, as to whether you know that *P*.

This, however, may seem to raise a worry. Surely, it is easy enough to raise the stakes. All I need to do is to offer you a bet on *P*: I'll give you a dollar (or ten dollars, or a hundred dollars – pick the smallest amount of money that you regard as being worth winning, given the bother of making a transaction) if *P* turns out to be true; you sign your house over to me if *P* turns out to be false.[13] But is it really true that that there are a great many things you regard yourself as knowing (the great many things that I am supposing you, with Williamson, regard yourself as knowing but on which you are not willing to bet your house), any one of which I can induce you to regard yourself as no longer knowing simply by offering you the foregoing bet?

The answer, I think, is "Yes." In offering you that bet, I would be offering you much the same bet as we imagined the holder of your mortgage has, in effect, offered you on the proposition that the bank will be open Saturday: "I'll give you rest and relaxation after a long day, if it's open," your mortgage holder has, in effect, told you. "You get your rest but give me your house if it isn't." We have already seen why it makes sense for you to withdraw your claim to know in the face of *that* offer.

The greater worry to most readers will, I think, be the fact that, on the present view, assessing what you know looks so epistemologically unimportant. It is a central mission of epistemology to tell us how, and in what way, we are to engage in the fundamental aim of inquiry: to assess, by purely epistemic lights, the merits of adopting doxastic attitudes toward propositions. We expect the adoption of these attitudes to answer to our desire to satisfy

scope of this sketch to explore what implications that the rationale I have offered (for your retracting your claim to know the bank will be open on Saturday) might have for an account of the conditions under which knowledge attributions are true. It is a thought about those conditions – the thought that facts about stakes do not, in general, affect what you know – that motivates the attractive thought about how you are to determine what you know.

13 Avishai Margalit actually issued a bet much like this to Isaac Levi in a review of Levi's *The Enterprise of Knowledge*. In that case, Margalit was proposing a bet on the truth of the conjunction of all the propositions Levi had at that point admitted into his corpus of knowledge – on which conjunction (as I noted earlier) Levi's account commits him to being willing to take any bet whatsoever.

our curiosity. We also expect the adoption of these attitudes to inform – in a manner that is neutral between different conceptions of the good – our rational decision making.[14] On the present view, it would seem that an epistemology could do all that without ever concerning itself with how you should decide what you know.

After all, the decision as to whether you know that *P* is, on this view, *no*t a purely epistemic one. It is, at least in part, a prudential one. Moreover, the decision to regard yourself as knowing that *P* is *not* neutral between different conceptions of the good. To regard yourself as knowing that *P*, you need to be (you open your position to criticism unless you are) willing to act as if *P*. Your willingness to act as if *P* will depend, in part, on how bad and how good you take the consequences of acting as if *P* to be. And that assessment will often reflect, and only reflect, *your* conception of the good.

But, again, I am not worried. These results were already in the cards. They were in the cards as soon as we rejected the attractive line of thought that, to determine whether you know that *P*, you need not in general concern yourself with what is at stake if you act as if *P*. Once we allow that deciding what you know requires you to advert to prudential considerations, we have conceded that the decision as to what you know cannot possibly play the fundamental epistemological role described two paragraphs back. Had we nothing to play that role, or were we able to see nothing of importance in our decisions as to what we know, that *would* be genuine worry. But neither of these unhappy possibilities actually obtains.

After all, if we are looking for doxastic attitudes the propriety of whose adoption is properly assessed by purely epistemic (and not at all by prudential) lights – doxastic attitudes whose adoption serves both the end of satisfying curiosity and the end of providing input into rational decision making that is neutral between competing conceptions of the good – they are ready to hand: states of confidence.[15] It is easy to see that the adoption of a state of confidence in *P* serves the end of satisfying curiosity.[16] And, thanks to the exponents

[14] The second expectation is more controversial than the first. I defend its propriety in "Decision Theory and Epistemology," in Paul Moser (ed.), *The Oxford Handbook of Epistemology* (New York: Oxford University Press, 2002), pp. 434–62.

[15] Levi calls them "credences."

[16] Or, at least, it is easy so long as one does not hold, with some advocates of Bayesianism, that you come (whether you realize it or not) already possessed of a precise degree of confidence assignment to every proposition you comprehend. On this view, the decision to adopt a state of confidence is not so easily construed as satisfying curiosity. There is never a case in which one adopts an opinion about *P* where one had no opinion before. But it seems to me that, both as a psychology and as an expression of a regulative ideal, this picture of your being endowed with a precise degree of confidence assignment is mistaken – as

of Bayesian decision theory, it is also easy to see how, in (to my mind) a very satisfying and intuitive way, our states of confidence can be marshaled to inform rational decision making. Finally, while some have complained that certain arguments Bayesians have offered for imposing a probabilistic coherence constraint on states of confidence rely unduly on nonepistemic considerations,[17] there can be no question but that the prudential consideration, which motivates you to abandon your knowledge claim in the bank case, provides no motivation whatsoever to change your state of confidence.

To see this last point, let us suppose that you are, in the bank case, very (but not maximally) confident – and warranted on purely epistemic grounds in being that confident – that the bank will be open on Saturday. Given this, and given the stakes as you initially take them to be, you decide not to go to the bank today. Then you learn that, if you persist in this decision and the bank is closed on Saturday, you will lose the house. It is sufficient reason not to be willing to act as if the bank will be open on Saturday – and so, given the fact about knowledge on which I have been insisting, sufficient reason to withdraw your claim to know it will be open on Saturday. But notice that this prudential consideration offers no reason whatsoever for you to change in any way how confident you are that the bank will be open. No amount of confidence in P short of the maximum possible commits you to acting, in all circumstances, as if P is true. In particular, when the costs of acting as if P when P is false are extremely high, it is eminently reasonable to be (indeed, unreasonable not to be) unwilling to act as if P, if you are only very (but not maximally) confident that P.[18] There is nothing at all odd about being very confident that P but, at the same time, being unwilling to bet your house on it.

So it is not your decisions as to what you know, but rather your decisions as to what states of confidence to adopt, that are meant to be made on purely epistemic grounds. And it is not your decisions as to what you know, but rather your decisions as to what states of confidence to adopt, that are meant (in a manner neutral between competing conceptions of the good) to provide the input into your decision making. But then what *are* your decisions as to

I explain in "Bayesianism without the Black Box," *Philosophy of Science* 56 (1989): 48–69. In this, I follow Levi's lead. See his "On Indeterminate Probabilities," *Journal of Philosophy* 71 (1974): 391–418.

[17] See, e.g., John Pollock, *Contemporary Theories of Knowledge* (Totowa, N.J.: Rowman and Littlefield, 1986), pp. 99–100; David Christensen, "Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers," *Journal of Philosophy* 93 (1996): 450–79, esp. 455–6. I argue that these charges are unfounded in *Decision Theory as Philosophy* (Cambridge: Cambridge University Press, 1996), pp. 40–3, and in "Decision Theory and Epistemology," sec. III.

[18] This is, of course, nicely captured by Bayesian decision theory.

what you know meant to do? Why should you care at all about, let alone take seriously the enterprise of deciding, what you know?

The answer, I want to suggest, is that to decide that you know that *P* is to decide (on pain of opening your position to criticism) to adopt a particular policy toward *P* – a policy of acting, in certain specific ways, as if *P*. It is not, let me emphasize, just a policy of acting as if *P* for the purpose of prudential (and moral) decision making. You might be happy to adopt such a policy toward *P* even if you hadn't the slightest idea whether *P* is true – provided that you were convinced that nothing of prudential or moral consequence hangs on whether *P* is true.[19] What more, then, is involved in the policy you adopt when you decide that you know that *P*?

It will help to remind ourselves of at least one of the conditions you will want to be satisfied before you decide that you know that *P*: that you can explain how you know that *P*.[20] By this, of course, I mean something perfectly colloquial: Nothing more or less than that you can provide what, for ordinary purposes would count as, an adequate answer to the question, "How do you know that *P*?"

In some cases, nothing very elaborate will be required to furnish an adequate answer. To "How do you know it's a Pakistani rug?" it might do to say, "By the way it feels" or "From its feel." In some cases (such as "How do you know this is a theorem?"), "It's just obvious" may do. Context influences what needs saying, what constitutes saying enough to say how one knows. But in some cases, the answer will take the form of something more like a proof: an answer in which you cite propositions to explain how you know that *P* – for example, when you explain how you know it is a Pakistani rug by appealing to its having a cotton warp.[21] And then it becomes important what you regard yourself as knowing. For you may not regard *Q* as available to cite by way of explaining how you know that *P* (you open your position to criticism if you regard *Q* as available to be so cited) if you regard yourself as

[19] Recall that you count as being willing to act as if *P* if there is no option that you acknowledge to be open to you that will have a consequence if *P* that no other option can better, yet you are unwilling to take that option. If, as far as you know, nothing of consequence hangs on *P* – that is, if, as far as you know, no option open to you has a consequence if *P* better than any other – this condition is satisfied.

[20] In contrast to Levi, I would maintain that if you regard yourself as knowing that *P*, you must (on pain of opening your position to criticism) regard yourself as able to explain how you know that *P* in the sense below. Levi holds (see note 10) that once you have admitted *P* into your corpus of knowledge, your being able to say how you know that *P* no longer is of any moment.

[21] Even in the first case, it can be argued, there is implicit appeal to the truth of a proposition – the proposition that the rug has a feel peculiar to Pakistani rugs.

not knowing that Q. Your explanation of how you know it is a Pakistani rug – "It has a cotton warp" – collapses if it can be legitimately replied, "But you don't know that!" So, part of the policy that you adopt when you decide that you know that P is a policy to treat P as a proposition available to be used to explain how you know other propositions.

Does what I have said in the last three paragraphs exhaust the policy implications of your deciding that you know that P? I very much doubt it. But I think that what I have said in those paragraphs goes a considerable distance toward making apparent what is at stake in deciding whether you do or do not know that P – and toward allowing us to see why it might make sense for you to care to make such a decision. What is at stake is whether to adopt the particular policy toward P just described – a particular policy for conducting inquiry and decision making. It is a policy that (like most policies) it will sometimes make sense to reconsider. You will want to reconsider the policy you have adopted when you acquire reason substantially to change how confident you are that P, or when the direction and/or size of the stakes in acting as if P undergo significant change. But, absent such conspicuous changes, the policy, if adopted, will issue you marching orders in a significant part of your life. It makes sense that you should care to decide whether to operate under such orders.

There is no shortage of evidence that we treat our decisions as to what we know as having just this sort of policy implication – and as important precisely for their having this policy implication. Consider one of the ways in which we use claims to knowledge. As Austin famously observed,[22] claims to know have illocutionary force. When you say to me, "I know that P," you give me your word that P. It opens you to criticism should I act on P and should P turn out be false – criticism to which you are not open if you say, instead, "I believe (I am quite sure) that P." Why is there this difference? Not because the latter assertion has no illocutionary force. When you say to me that you believe (or that you are quite sure) that P, you often mean that I should believe (be quite sure) too. The difference lies in the fact that, in saying you *know*, you are not telling me that you have a certain opinion, and meaning that I should join you in that opinion; you are telling me that you are prepared to act as if P in the manner described earlier – and that you mean me to be prepared to do so as well. (So, if I act as if P and things go badly because P turns out to be false, I can criticize you for having meant me to have so acted.) We are often in the

[22] See J. L. Austin, "Other Minds," in Austin, J. O. Urmson, and G. J. Warnock (eds.), *Philosophical Papers*, 3rd ed. (Oxford: Oxford University Press, 1979), pp. 76–116, 99–101.

market for something that will tell us how to act. Avowals of knowledge meet that demand, where expressions of confidence do not.[23]

Moreover, they do so in just the way the foregoing discussion suggests that they would. If I suspect that I have much more important things at stake in acting as if *P* than you have or are aware of my having, I may be happy to join in being as confident that *P* as you say you are, but refuse to treat you as knowing that *P*.[24] Likewise, if you come to realize that I have much more important things at stake in acting as if *P* than you have hitherto had (I will be risking my house or my life), you will be much more reluctant say to me, "I know that *P*" than you were before.[25] But in the other cases, in which we perceive our stakes in acting as if *P* to be comparable (and, of course, we take ourselves and others to be responsible epistemic and prudential decision makers), we will find the trade in frank knowledge avowals (and so our respective decisions as to whether we know that *P*) useful and helpful – useful and helpful in a way that the trade in expressions of confidence (and so our respective decisions as to how confident we ought to be that *P*) is not.

IV

A final note. I promised at the end of section II to sketch a way of thinking about knowledge that is compatible with the abandonment of the attractive thought that your decision as to whether you know that *P* should be unaffected by what you take to be at stake in acting as if *P*. What I did not promise, but delivered nonetheless, is a way of thinking about knowledge that is also fully compatible within an otherwise unadulterated Bayesian probabilism – a way

[23] My thinking about the importance of knowledge avowals has very much been influenced by conversations with Adam Leite.

[24] Just as I may not regard myself as knowing that *P* yet, at the same time, be unwilling to act as if *P*, so I may not regard *you* as knowing that *P* and yet, at the same time, be unwilling to act as if *P*. Once I grant that *P* is known (whether by me or anyone else), I open my position to criticism if I am unwilling to act as if *P*.

[25] But not necessarily more reluctant to regard yourself as knowing. If you take my cares as your own, and my view of the stakes and decisions confronting me as your own, then the reluctance will extend to what you regard yourself as knowing. Otherwise, the reluctance will simply be a reluctance to *say to me* that you know that *P*. It will be nothing more or less than a reluctance to issue advice as to how to act to someone who, you are confident, is apt to use the advice in a way that you would not want to see it used (e.g., in a way inadvisable for someone with his or her cares). The willingness to act as if *P* associated with knowing that *P* does not include a blanket willingness to say that you know that *P*. Whether you are willing to say that you know that *P* may be affected by prudential and moral considerations over and above what hangs on *P*'s truth value. It is entirely compatible with your acting as if you are carrying contraband in the trunk of your car that you say to the border guard that you know you are carrying no such thing.

of thinking about knowledge that can simply be appended to, and in no way affects the workings of, the Bayesian probabilist picture of one's choice. My conceit is that my sketch has provided the outlines of a compelling case to the effect that there is a better way within the Bayesian framework to think about our ordinary practice of attributing knowledge to ourselves than there is available without.

But, it might be complained, it is nothing more than a conceit. My account (it might be complained) leaves completely unanswered crucial questions – the very questions that Levi's account of knowledge was designed to answer. The preceding pages are full of talk of what happens when you *learn* that the stakes in the bank case change; of what is the case should you *acknowledge* that there is an option open to you that has a certain character. How are we to understand such talk? Levi has an answer: It is talk about the introduction of propositions into a corpus of knowledge of precisely the sort Levi's account describes. What answer do I – with only my unadulterated Bayesian probabilism and my (from the point of view of the workings of that probabilism) epiphenomenal account of the import of decisions as to what one knows – have to offer in its stead?

It seems to me that I have everything I need. The complaint is this far right: An unadulterated Bayesian probabilism offers only conditional advice. It tells you that *if* you face a decision problem with features x, y, and z, *then* here is how your preferences over its options are constrained. It doesn't tell you when you are warranted in regarding yourself as actually confronting a decision problem that has features x, y, and z. It doesn't tell you under what circumstances its advice is actually applicable.

But notice that, to apply the advice issued by an unadulterated Bayesian probabilism, you do not need to get into a state in which you are willing to bet all and everything that you are facing a decision problem that has features x, y, and z. Think of the conditional advice posted next to elevators: "In case of fire, use stairs." You do not need get into a state in which you are prepared to bet all and everything on the hypothesis that there is a fire before you can apply the advice posted. You need only get into a state in which you are willing, in your present circumstances, to act – in precisely the respects I have described above – as if there is a fire. So it is with the conditional advice issued by an unadulterated Bayesian probabilism: To apply it, you need only get into a state in which you are willing, in your present circumstances, to act (in those respects) as if you are confronted by a decision problem with features x, y, and z.

Does a regress threaten? It might be thought that it does. Let Q be "You are confronted with a decision problem with features x, y, and z." It might be thought that, in order to be willing to act in your present circumstances as if *Q*, you need have to determined that there is not, among the options open to you, one that will have a consequence if *Q* that no other option can better, yet you are unwilling to take that option. That will, in turn, require you to be willing to act as if there is no such option, which will require you to have made yet another determination, and so on, *ad infinitum*.

But the thought goes awry. Your being willing to act as if *Q* does not require you to have made a determination of a certain sort. It requires only you *not to have made* a determination of a certain sort: not to have determined that there is, among the options open to you, one that will have a consequence if *Q* that no other option can better, yet you are unwilling to take that option. Entering into the state in which you are willing to act as if you are confronted by a decision problem with features *x*, *y*, and *z* thus does not require you to have made an infinite number of determinations. It is entirely compatible with the forgoing that there are circumstances in which warrantedly being in (or even entering into) that state may require very little of you.[26]

What, then, would *be* required of you before you are warranted in entering into the state of being willing to act as if *Q*? It is a question whose answer would take me well beyond the scope of the sketch I have been offering here. But this much, I hope, is clear: If what I have just said is right, it is precisely the sort of question that an account of how to decide what you know – understood in the way in which I have been arguing we should understand it – would answer.

[26] Compare the sort of modest foundationalism with respect to knowledge that, for example, Austin championed: You count as knowing a proposition just if you can meet every legitimate challenge to the propriety of your claiming to know that proposition (i.e., every legitimate "How do you know that?" question). Austin held that what challenges were legitimate was very much influenced by context. And he held that, in some contexts, a claim to knowledge was open to no legitimate challenge: There is nothing you have to be able to do in order to earn your entitlement to claim knowledge. Levi's own pragmatic epistemology is committed to the view you are entitled to your current corpus of knowledge without having now to do anything to earn that entitlement. See again the passage I quoted in note 10.

16

Levi's Ideals

Sven Ove Hansson

1. INTRODUCTION

Isaac Levi's work has had what seems to be a lasting impact on several fields of philosophical inquiry. I am myself one of the many philosophers on whom he has had a decisive influence. His book *Gambling with Truth* was one of the major inspirations that led me to study philosophy, and I have continued to be an eager reader of his books and articles.

In this contribution I focus on Isaac Levi's choice of formal structures for the representation of beliefs. This aspect of his work exemplifies his influence; a large group of reseachers in decision theory and formal epistemology have either followed his proposals or taken them as starting points for their own developments. Since philosophy proceeds by criticism rather than by praise, I focus on what I perceive as possible problems and needs for clarification.

2. TWO TYPES OF IDEALIZATION

The representation of philosophical subject matter in formal language is always the outcome of an idealization. There are two types of idealization that should be carefully distinguished between, although they are often intertwined. First, to idealize can mean to simplify for the sake of clarity. The resulting formal model is an ideal in the sense of "[s]omething existing only as a mental conception" (Oxford English Dictionary). To idealize in this sense means to perform a "deliberate simplifying of something complicated (a situation, a concept, etc.) with a view to achieving at least a partial understanding of that thing. It may involve a distortion of the original or it can simply mean a leaving aside of some components in a complex in order to focus the better on the remaining ones." (McMullin 1985, p. 248). Second, to idealize can mean to formulate an ideal in the sense of something that is "perfect or supremely excellent in its kind" (OED). In the present context, this means that the formal model represents perfectly rational patterns of belief or belief change.

Like most other models of beliefs and decisions, Levi's formal framework is idealized in both these senses. In other words, it is both (1) idealizing-simplifying, that is, it leaves out many of the complexities of real life, and (2) idealizing-perfecting, that is, it represents patterns that satisfy higher standards of rationality than what actual doxastic agents do in real life.

In my view, there can be no doubt that both of these processes of idealization are indispensable in formal epistemology. Actual processes of belief change are so complex that substantial simplifications are necessary in order to obtain a model that is at all manageable. In other words, certain properties of real-world doxastic behavior have to be omitted in order to make others come out more clearly. Furthermore, in order to be able to discuss requirements of rationality, we need to develop models in which such requirements are adhered to more consistently than what we do ourselves in our daily lives. In other words, we need models of belief that are idealizing-perfecting as well as idealizing-simplifying.

3. DOXASTIC COMMITMENTS

According to Isaac Levi, the deliberating agent is simultaneously in three states, namely, "a state of full belief," "a state of credal probability judgment," and "a state of value commitment" (Levi 1997, p. 5).[1] Clearly, the isolation of each of these three specialized states from the overall state of mind is an idealization in both senses. The separation of (full and probabilistic) beliefs from value components does not mirror how human minds work, but arguably it mirrors how they would work if we could draw the fact-value distinction with perfect precision.

Each of the three states is assumed to satisfy certain rationality requirements. The state of full belief is assumed to be consistent and closed under logical consequence. Changes in that state are assumed to satisfy a series of conditions for rational belief change (Levi 1991, 1996). The state of probabilistic judgments is assumed to satisfy the calculus of probabilities, which

[1] The state of full belief and that of probabilistic belief differ in their approaches to degrees of belief. There are two notions of degree of belief. One of these is the static concept of the *degree of confidence* with which a belief is held. The other notion is the dynamic concept of *degree of resistance to change*, that is, how difficult it is to change the belief (Hansson 2003). In standard probability theory, and consequently in models of probabilistic belief, no distinction is made between these two notions of degree of belief. They are both represented by the probability function. In contrast, states of full belief do not conflate them. With respect to confidence, states of full belief are dichotomous, that is, a sentence is either fully believed or not believed at all. However, most models of states of full belief allow for different degrees of resistance to change (Levi 1991, p. 141).

specifies both its static and its dynamic properties. The value state is assumed to satisfy requirements of rational preferences, including transitivity.

For each of these sets of rationality requirements, there is an extensive literature showing both that ordinary human agents often disobey them and that they may be perfectly rational in doing so. Logical closure of the state of full belief is a clear example of this. At first sight, it may seem sensible to require of a rational agent that he or she believes in all the logical consequences of his or her own beliefs. However, this implies that he or she believes in all logically true sentences, including all mathematical theorems. Such logico-mathematical omniscience is of course far beyond human capabilities.[2]

Isaac Levi makes sense of this large deviation from our actual doxastic behavior by assigning to the state of full belief the role of representing not what an agent actually believes but what he or she is committed to believe (Levi 1977, 1997, 2002). Although we are unable to draw all the conclusions that follow from our beliefs, we can nevertheless be doxastically committed to believe in them.

It has to be recognized, however, that this interpretation involves a large deviation from the common understanding of what it means to be committed to something. In ordinary parlance, commitment is subject to a "committed implies can" restriction that parallels the "ought implies can" restriction. To the extent that I am committed to believe in exactly those mathematical statements that are true, this is a commitment in a sense entirely different from that in which I am committed to keep my promises and repay my loans.

The philosophical relevance of results obtained in models with a high level of idealization-perfection depends on the chosen purpose of philosophical inquiry. On one view, rationality per se is a meaningful subject matter for philosophy. Then the reasoning powers ascribed to ideal rational agents need not be restricted by what is humanly – or even physically – possible. We can then follow Levi in focusing on ideal agents with transfinite cognitive capacity.[3] According to another view, philosophy's subject matter centers around human beings and the conditions under which they live. According to the latter standpoint, we should focus our interest on another type of ideal

[2] Due to the mathematical convenience of logical closure, it can be the outcome of idealization-simplification (but not of idealization-perfection) in models attempting to reflect the properties of finite minds. The important difference is that in this case, results depending on logical closure are treated as anomalies caused by the imperfections of the model.

[3] This approach is problematic from a methodological point of view. It is not clear how we can be able to choose principles of rationality for agents with reasoning powers that are qualitatively higher than what we have ourselves.

agents, namely, agents that have limited cognitive capacity of which they make rational use.

4. NORMATIVE RELEVANCE

Levi fully recognizes that his requirements on the three idealized substates of mind represent a far-reaching idealization-perfection. He readily admits that we do not come anywhere near to such a state. Nevertheless, he claims that models involving such states can be used for the purpose of finding out how one ought to reason.

> In order to achieve a state of rational equilibrium, X would need unbounded memory capacity and computational resources as well as abilities for self-knowledge that few come close to possessing. . . . The fact that we do not even come close to satisfying requirements of rationality precludes the serious usefulness of principles of rationality as explanatory, predictive, or descriptive principles. They are not even useful idealizations of human behaviour for explanatory and predictive purposes. States of rational equilibrium are, however, very important normatively. (Levi 2002, p. 217)

However, it is far from self-evident that principles developed for highly idealized reasoners provide reliable normative guidance for agents with much more limited reasoning capabilities. The best use of limited cognitive resources may require that one follows principles and processes that would not be useful for logically omniscient beings.

To clarify this, we can compare with the role of ideal worlds in deontic logic. An ideal world is a hypothetical state of the world in which the behavior of all agents is morally perfect. In standard deontic logic, ideal worlds are used to determine what is morally best to do, according to the recipe that you should do exactly what you would do in all possible worlds (Føllesdal and Hilpinen 1970). It can easily be shown that this will lead us wrong. To take just one example, in an ideal world there is no racism and no sexism, and therefore there is no one who fights racism or sexism. If we judge our obligations by what we would do in the ideal worlds, we should not fight any of these atrocities. This is a counterintuitive result from which we should conclude that ideal worlds are not well suited for action guidance (Hansson 2001, pp. 140–1).

For formal epistemology it is a crucial issue whether or not the use of models with unrealistically high levels of cognitive perfection give rise to similar problems. In particular, can patterns of reasoning that have been derived from models of logically omniscient beings be unsuitable for action guidance in the real world because they do not take into account factors that the nonideal, contrary to the ideal, reasoner has to deal with? My answer to this question is yes, and I substantiate this answer below with an example.

5. INFINITE MINDS AND FULL BELIEFS

According to the Bayesian ideal of rationality, a rational subject should assign a definite probability value to each statement about the world. Only logically true sentences are assigned probability 1. Nonlogical propositions can, at most, be assigned high probabilities that are marginally lower than 1. The resulting belief system is a complex web of interconnected probability statements, of which all that do not represent logical truth have probabilities below 1 (Jeffrey 1956). Although sensible arguments can be given for this approach, it has one decisive disadvantage: In practice, such a belief system would be unmanageable for human subjects (McLaughlin 1970). Our cognitive limitations are so severe that massive reductions from high probability to full belief (certainty) are indispensable in order to make us capable of reaching conclusions and making decisions. In other words, we have to treat a large number of contingent statements provisionally in the same way as if they were known with certainty to be true. This reduction to full belief, or "fixation of belief" (Peirce 1934; Levi 1991), helps us to achieve a cognitively manageable representation of the world. For this reason, models based on states of full belief can represent certain features of our doxastic behavior more realistically than probabilistic models.

However, this argument for models including states of full belief is applicable only to agents with limited cognitive capacities. An (ideal) agent with unlimited reasoning powers has no reason to reduce high probabilities provisionally to full beliefs, since he or she can reason conveniently with unreduced probabilities.

In the choice between levels of idealization, Levi seems to want to have it both ways. On the one hand, he advocates a high level of idealization on which agents have transhuman, even transfinite, reasoning abilities. On the other hand, he assigns states of full belief to these agents. As we have seen, a state of full belief is easily motivated on a level of idealization that aims at representing reasoning with cognitive limitations, but not on a level that represents logical omniscience.

Most formal representations of reasoning contain either a model of full belief or one of probabilistic judgments, but not both. How these two states can coexist and interact is one of the most important unsolved problems in formal epistemology. To make progress here, it seems essential to take the reasoning agent's cognitive limitations into account.

6. AN ISSUE OF DETERMINISM?

In a recent article, Levi connects his account of belief states as representing commitments to an attempt to avoid a determinism problem in formal epistemology. He is worried that belief states representing the inquirer's actual doxastic behavior will have to be "dispositional states (or biological states of some kind)." Such states are determined and cannot be changed "by choice." In contrast, he says, "[c]hanges in commitment can plausibly be subject to the agent's direct control" (Levi 2002, pp. 211 and 216).

It is debatable whether the commitment maneuver solves the determinism problem. If the agent's doxastic commitments consist in believing the logical consequences of his or her actual beliefs, then his or her commitments are biologically (etc.) determined to the same extent as his or her actual beliefs are biologically (etc.) determined. More important, issues of rationality can be treated independently of issues of determination or causation. This can be seen from the use of models of belief change to describe data-base updating or other operations performed in a computer. We can discuss whether or not the operations performed by a machine satisfy various criteria of rational belief change, just as we can discuss the corresponding questions for operations performed by a human being. Rational behavior does not require nondetermination or autonomous control.

REFERENCES

Føllesdal, Dagfinn, and Risto Hilpinen. 1970. "Deontic Logic: An Introduction." In Risto Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings*, pp. 1–35. Dordrecht: Reidel.

Hansson, Sven Ove. 2001. *The Structure of Values and Norms*. Cambridge: Cambridge University Press.

Hansson, Sven Ove. 2003. "Ten Philosophical Problems in Belief Revision." *Journal of Logic and Computation* 13: 37–49.

Jeffrey, R. C. 1956. "Valuation and Acceptance of Scientific Hypotheses." *Philosophy of Science* 23: 237–49.

Levi, Isaac. 1967. *Gambling with Truth: An Essay on Induction and the Aims of Science*. New York: Borzoi.

Levi, Isaac. 1977. "Subjunctives, Dispositions and Chances." *Synthese* 34: 423–55.

Levi, Isaac. 1991. *The Fixation of Belief and Its Undoing. Changing Beliefs through Inquiry*. Cambridge: Cambridge University Press.

Levi, Isaac. 1996. *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference, and Nonmonotonic Reasoning*. Cambridge: Cambridge University Press.

Levi, Isaac. 1997. *The Covenant of Reason: Rationality and the Commitments of Thought*. Cambridge: Cambridge University Press.

Levi, Isaac. 2002. "Commitment and Change of View." In José Luis Bermúdez and Alan Millar (eds.), *Reason and Nature: Essays in the Theory of Rationality*, pp. 209–31. Oxford: Clarendon.

McLaughlin, A. 1970. "Science, Reason and Value." *Theory and Decision* 1: 121–37.

McMullin, Ernan. 1985. "Galilean Idealization." *Studies in History and Philosophy of Science* 16: 247–73.

Peirce, Charles Sanders. 1934. "The Fixation of Belief." In Charles Hartshorne and Paul Weiss (eds.), *Collected Papers of Charles Sanders Peirce,* vol. 5: *Pragmatism and Pragmaticism*, pp. 223–47. Cambridge, Mass.: Harvard University Press.

17

The Mind We do Not Change

Wolfram Hinzen

1. INTRODUCTION

For many years Isaac Levi has been a staunch defender of a strictly normative and prescriptive conception of rationality. The origin and motivation for this crucial commitment, as it transpires particularly clearly in *The Covenant of Reason* (Levi 1997; henceforth CR), has been Levi's exploration and development of the Peirce-Dewey "belief-doubt" model of inquiry. On the latter, justifiable change in state of belief is a species of rational decision making. This is what motivates Levi's concern with "rationality" in the first place (CR, 20). In fact, no substantive commitment on what rationality substantively *is* – or on what it is to be rational – emerges from this theoretical interest. In particular, we are not told what beliefs or values we should have, which ones it is rational to have, or how we should base our beliefs on "evidence." Rather, principles of rationality are primarily justified instrumentally through their *regulative use* as formal constraints on well-conducted inquiry and problem solving, no matter the domain, be it science, politics, economics, technology, or art, or even simply the personal decisions we face in daily life. Given their exceeding generality, we can only expect constraints on the coherence of choice to be both formal and weak. Principles of rationality are to be kept immune from revision if a general theory of how rational changes in point of view are to be justified is to be possible at all (CR, 24). But I understand this is to be an essentially practical necessity, which does not depend on a notion of what the "essence" of rationality is. We are dealing with a fundamentally *instrumental* conception of rationality here (CR, 16), not with a conception in which rationality is something to strive for or to analyze for its own sake.

Dedicated to Isaac Levi, who changed my mind completely during memorable years in New York. But it changed again – in ways this chapter documents. We should be looking for something stable!

I find much to admire in this vision, whose at times quite radical minimalism and modesty as regards the study of rationality contrasts quite sharply with more portentous conceptions of it (and of us as "essentially rational beings"): For example, it offers little support for the idea that the theory of rationality can be appealed to in an effort to explain and "rationalize" the political and economic organization of modern societies, say as the forming of a "contract" between naturally constituted rational individuals confronting each other as competitors for scarce resources in a state of nature. On a different score, and despite its decidedly narrow focus, Levi's vision of rationality has clear and ramified implications for the agenda of twentieth-century philosophy, not only with regard to metaphysical issues of correspondence and reference or the nature of propositions (cf. Levi 1991), but also with regard to the issue of meaning, the analytic, and the a priori. The best parts of the latter, one might argue, fall out from an account of how our revisions of belief are constrained (so that analytic truths, in particular, would be an epiphenomenon of the fact that beliefs have varying degrees of entrenchment).

All that said, I use this opportunity to take a step back and read Levi somewhat against himself, confronting his vision of philosophy and rationality with another, more naturalistic one, in ways that may not only illuminate it, but also change it internally. Particularly if rationality is fundamentally instrumental, *naturalizing* rationality seems an option, contrary to what Levi suggests.[1] There is, I emphasize, no question that metaphysical issues such as naturalization are peripheral to Levi's main concerns. Even epistemological issues have an unclear status, if these, say, include debates over the correctness of empiricist versus pragmatist or rationalist so-called theories of knowledge. Levi, while of course a committed pragmatist, does not actually give us a "theory of knowledge," especially if this includes a conceptual definition of what knowledge is (cf. Levi 1980, henceforth EK, sec. 1.9). An analysis of the "Enterprise of Knowledge" – a theory of justified *change* of belief or states of knowledge – is a quite different enterprise. Still, I argue that discussing features of both naturalism and rationalism helps to bring important features of Levi's philosophy into clearer view. The bottom line is that while it is true, of course, in one sense, that we "change our minds" (how we should do so being Levi's lifelong theme), there is also the mind we do not

[1] Probably *only if* this is so. I see little scope for a naturalization of intrinsic (rather than instrumental) value, and Dennett's (1995, chs. 16–17) "naturalized ethics" as a failure in this respect.

change: the rational mind we happen to have, by virtue of our evolution and nature.

2. LEVI'S NORMATIVISM

On Levi's view, the theory of rationality is not in the service of telling us something about the natural world. Rather than being a descriptive theory concerned with what is *true*, it sets a *standard* that may guide the deliberating agent in the monitoring of his or her own decisions and changes of belief. It has neither explanatory nor predictive uses. Since there is no limit to human confusion, we will typically not live up to this standard. In this very sense, being rational cannot be a property that empirically characterizes us as humans. As is more generally recognized in discussions of bounded rationality in Bayesianism, poor memory, limited computational capacities within given time constraints, acting on conflicting motives without resolving the conflicts inherent in them, emotional stress, lack of self-knowledge, issues of identity and religious value not reducible to material consumptions, and so on will necessarily lead us to act "irrationally" in the light of received views of rational choice in classical economic and Bayesian economic theory. The disconfirmation of rational choice theory understood as an empirical one appears to be one of the truly robust results achieved in the human sciences (Conlisk 1996).

Thus it cannot be rationality that makes us agents. Levi's solution to this problem for classical rational choice theory is that we are not the "rational being" but the being that *tries* to be rational, setting itself and using a standard of rationality for evaluating its own actions (cf. CR, 6–7).

Levi's normativism ties in naturally and clearly with an explicit verdict against naturalism, in that something is rational only if you *evaluate* it according to a certain standard, and there is no road leading from facts to values (CR, 14). The *prescriptive* does not reduce to the *physical*. No natural being *can* as such be rational. Since Levi effectively and interestingly equates the mental or intentional with the prescriptive – though he makes the point almost in passing (CR, 14) – the mind never comes into view as a natural object with its own structure and function. It rather only *arises* from the evaluation of certain natural facts, such as utterances and actions. In evaluating them, they come to be seen as either *generating* certain commitments – for example, by saying what he or she does, the agent becomes committed to be disposed to endorse certain consequences – or as *fulfilling* them – thus I may evaluate the handing-over of an amount of money as the fulfilment of a promise or contract. Drawing up agreements for a contract, in Levi's terms, (necessarily) involves events that are describable in purely physical terms, such as the putting-down of a pen on paper.[2] But using a standard of rationality, we can redescribe these very events as generating an obligation, a redescription in which prescriptive moral principles and laws of contract are necessarily involved.

Note that it is not *quite* the case, here, that there is nothing to say about the mind in an empirical and naturalistic (or "psychological") perspective. On the contrary, human rational behavior is subject to empirical study. Crucially, however, we can study empirically only what we evaluate normatively too. There can be no such question as whether or how rationally an agent behaved, without a view on what it would have been for that agent to have behaved rationally in this circumstance – that is, without viewing his actual actions as a *performance* realizing certain attitudinal *commitments*. With the view on the agent's commitments in place, his degree of fulfilment of them through his actions becomes an empirical question. Commitment and performance are correlative on this view, showing that there is no empirical dimension to the study of human action and mentality without a normative one. As Levi puts it in a somewhat Kantian mood (CR, 16), commitment without performance is "empty" – the rational agent must be viewable as attempting to realize his commitments, even if he does not live up to them – but performance without commitment is "blind" – for understanding of an agent's actions comes when we can see his actions as attempts to fulfil his commitments.[3]

Levi's strict normativism is premised on the *unbounded* nature of rationality, that is, its nonrelativity to the limits of our capacity to fulfil rational obligations we incur. It is the unbounded nature of rationality that makes us rational failures and deprives the theory of rationality of an empirical (explanatory and predictive) content.[4] A more provocative way of putting this same insight is that the empirical study of the human mind reduces to the *clinical* study of its rational deficiency. No theoretical insight into the structure of rationality

[2] The example is Levi's. I myself would want to note here that even terms such as "pen" and "paper" have no status in physical theory, hence are also not describable in physical terms alone.

[3] Although Levi's reference to Kant is only in passing, there is, I believe, a much more systematic connection here between Levi's non- or antinaturalism and a Kantian "two-worlds" view. Much of his emphasis on conflicts between a first-person and a third-person perspective on choice or between the autonomy of the agent and the predictability of its trajectory viewed as a naturalistic entity (see section 6) relates to a Kantian mode of thought (cf. also Rabinowicz 2002, pp. 91–2).

[4] "Principles [of rational belief, desire, and choice] fail to contribute to explanation of behaviour as physically described because such principles are false as applied to human beings" (CR, 7 n. 2; cf. 36–7).

flows from this. A *practical* need arises, rather, to improve, through therapy, training, and prosthetic devices, the agent's performance against the background of his commitments. Much in Levi's philosophical frame revolves around the distinction between *inquiry* – called for when a change of belief is to be justified – and *therapy* – called for when an agent does not realize which dispositions he is committed to have (CR, 11).

It is the verdict against bounded rationality, however, that I would take least issue with here. I find it plausible that if we clearly see what a rational solution to a decision problem is, and still do not decide in practice in accordance with this insight, there is no reason here to lower our standards. As long as a therapy can be sought that brings us closer to the rational ideal, I do not see what should lead us to abandon the ideal *as such*. Suppose indeed we lower the standards. Then,

> no matter how we trim our principles of rationality, there will always be predicaments so complex and stressful as to preclude the applicability of the eviscerated standard. Evisceration will continue until nothing of interest is left to carve out. (CR, 8)

In addition to this compelling argument (in my view), which makes bounded rationality approaches self-refuting, one might reason as follows. Arguably, modified versions of classical rationality designed to cope with anomalies blurring the predictive power of classical rationality tend to retain a basic orientation to the classical conception of full rationality. The theorist of bounded rationality will typically test which assumptions lead to behavior different from what unbounded rationality predicts. In this sense, as Foley (2003) remarks,

> bounded rationality is an epicyclic extension of rationality theory, and struggles to define itself in terms other than its deviations from the results of full substantive rationality. (p. 4)

Given the heterogeneity of causes for anomalies it seems unclear how a unified theory of bounded rationality can be achieved, "which instead tends to degenerate from an explanatory framework into a descriptive language" (ibid.).

Levi's framework, by embarking on strict normativism, avoids these pitfalls. This becomes clear already in the decision to model beliefs as *commitments*, for commitments are by their nature not inconsistent with a failure to live up to them (a broken promise is still a promise, after all). Rather than concluding that the classical theory of rationality is simply a disconfirmed theory, it lifts its laws to the level of regulatory principles for self-criticism that have a normative status only. Rational beliefs can now remain structured according to the very strong logic involving the condition $B(A) \rightarrow A$: whatever the agent fully believes is true, actually is. Doxastic coherence demands this, Levi argues (CR, 66). The one thing that saves us from absurdity here is to give up viewing that logic of belief as a logic characterizing a set of truths (hence as describing the real world). The logic of this kind of coherence, in terms adapted from Ramsey, is the "logic of consistency," not the "logic of truth" (CR, 44–5).

But while there is nothing to object to the beauty and ingeniousness of this solution, it also leaves us in the dark about what an explanatory framework for the explanation of human action should be. Figuring out which therapeutic treatments would do best for a given agent struggling with rationality is not a task catching everyone's ambition. And should we content ourselves that there is little scope for a naturalistic inquiry into the rational structure of the human mind at all? Or that, indeed, there is no specifically "human" mind to speak of, given that, on Levi's view, a normative redescription of physical events according to standards of rationality can be provided for the actions of more abstract agents, such as institutions and governments, as well?

3. NATURALISM AND RATIONALISM

Levi's picture of the mind could hardly be in greater conflict with the rationalist tradition, as I have myself construed it in my own ways in a forthcoming book (Hinzen forthcoming). On this alternative view, man's reason is an expression of his nature, and to the extent that man's rational mind can bring light into the structure of the real, reality itself is rationally structured. Human reason is subject to empirical and naturalistic study – our minds are structured by grammatical, moral, aesthetic, and other forms of knowledge – and in pursuing this study there is no intrinsic need to hold human performance against a prescriptive norm.[5] The human mind and its intrinsic structure is as such what is rational in nature, and its rationality has no more to do with how we evaluate nature than other aspects of the natural world. Stipulated norms

[5] Quite the contrary, I tend to agree with Wertheimer's (and Goodman's) "Factunorm" principle, according to which how we do think is necessarily evidence for the principles of rationality. "We are (implicitly) accepting the Factunorm Principle whenever we try to determine what or how we ought to think. For we must, in that very attempt, think" (Wertheimer 1974, pp. 110–1).

on human behavior and public policies deriving from contingent values have to be held against human nature as a constraint on which norms and policies we should have. The science of morality is, as on the Humean (1978) conception, part of the "science of human nature," a study of an aspect of human beings qua natural beings and an inherent part of the new sciences of his day. Moral psychology becomes the empirical study of our species-specific faculty of reasoning and moral judgment, cartographing its structure and content.[6]

Although outmoded in philosophy, the scientific basis of this picture in our own time seems clear enough.[7] Cognitive science has identified and described structures inherent in the human mind that enable our various cognitive competences to develop in the uniform way they do (Hirschfeld and Gelman 1994; Mehler and Dupoux 1994; Bjorklund and Pellegrini 2002; Pinker 2003). Naturally, both the methodologies and the results of these works are open to revision and empirical refutation. What matters is they require empirical counterarguments, whereas claims that human nature does not exist, which abound in the philosophical literature (e.g., Rorty 1990, evoking and echoing a long tradition of continental philosophy including Heidegger, Gadamer, and Derrida), are usually not argued for on empirical grounds and are often simply taken for granted.[8]

Human nature, for all empirical analyses suggest, exists in the sense of intrinsic and species-specific structures characterizing the human mind. While the human language faculty is a classical example, more recent studies suggest that infants have an innate expectation regarding the nature of objects (objects move as bounded wholes, they are solid, they move on connected paths, they may be inanimate or animate, etc.). A recent study suggests that five-month-old infants do not readily apply physical principles to humans that they apply

[6] Harman (1998) suggests just such an agenda for moral theory.

[7] The scientific basis for the Quinean project of a "naturalization of philosophy," for which Skinner's radical empiricism was instrumental, remains unclear in comparison. From the viewpoint of traditional rationalist or Chomskyan rationalism, Quinean "naturalism" is the opposite of naturalism in the Humean sense (Hinzen forthcoming).

[8] In the analytic tradition, equally, it seems that no notion of human nature has ever centrally figured. Frege and Wittgenstein drew philosophers' attention away from natural language as an intrinsic property of the human mind. For Carnap, the only exception to the basic correctness of (a pragmatist version of) empiricism was our knowledge of logic. Quine considered the attempt to study, on an empirical basis, the innate structures of the mind that enter into human language use a form of "folly." It seems no exaggeration that the mind as the rationalist – in Quine's case, it was Chomsky's reincarnation of it – proposed to study it played virtually no role in the philosophical reflection on the mind ever since Frege, Russell, Wittgenstein, Carnap, and Quine set twenteeth-century philosophy on its course.

to physical and inanimate objects, and in fact have trouble viewing them as material objects at all. Overall, the study concludes,

> young infants may have different modes of construal for humans versus inanimate objects: humans are construed in terms of social and intentional actions, while inanimate objects are interpreted via a system sensitive to object physics.
>
> (Kuhlmeier, Bloom and Wynn 2004)

Finally, construing actions as social includes, according to some views, construing them as normative (or as subject to contractual obligations, as in Cosmides and Tooby's classical 1992 analysis). To the extent that these empirical conclusions are sensible, one explicit motive for Levi's antinaturalism disappears. As noted, Levi takes Brentano's thesis of the irreducibility of the intentional to the physical on board, collapsing it with the fact-value dichotomy. Somehow, the idea seems to be, naturalistic science can deal with what it is to put a pen down on paper, but it cannot deal with what it is to take up an obligation (as by signing a contract). But, from the viewpoint of human beings (infants, indeed), the world is not first or primarily "physical," and then the intentional comes as an additional element in virtue of these humans adopting certain values for redescribing nature. For all developmental psychology suggests, the intentional does not have to do with contingent *values* at all. It has to do with what creatures we are, and what internal structures we bring, unconsciously, to bear on the reality we happen to be embedded in. We structure our environment into both intentional and nonintentional ingredients, but the former are not less "natural" for that nor in need of "naturalization." The physical-intentional distinction has first of all a cognitive basis, not a metaphysical one, and in this sense we do not need normativity for there to be mental aspects in the things that surround us. We need a mind of the right sort. While that account questions the very basis of the naturalization project (which depends on setting up a *metaphysical* dichotomy), it also questions the idea of antinaturalism, which depends on the same spurious dichotomies.

If mental and intentional aspects of nature (beliefs, desires, intentions, etc.) are simply there in our environments because we have the kinds of minds we have, the metaphysical idea of a "reduction" becomes neither affirmable nor deniable. It becomes simply obscure. Making the existence of beliefs and desires contingently dependent on evaluating them according to certain norms is no more invited here. What we are left with is a *methodological* (rather than metaphysical) naturalism that inquires into the mental aspects of animate beings in no other ways than it inquires into their physical aspects. Studying human language is to study just another organic system with its internal structure and function, like the immune system or the circulatory system.

Neither pragmatism nor rationalism, I take it, is generally speaking inconsistent with a methodological monism and naturalism in this sense. Pragmatism wasn't in the case of Dewey, an issue to which I turn shortly. Rationalism certainly wasn't either, not, for example, in the case of Descartes, whose supposed "dualism," to the extent that this term makes sense (cf. Baker and Morris 2002), was formulated as a part of the natural sciences of his day. It certainly isn't either in the case of Chomsky's "rationalist psychology" (cf. Chomsky 2002) or the biolinguistic tradition (Lenneberg 1967).

Pragmatism met with naturalism particularly in the Dewey of *Experience and Nature* (1926).[9] Both Levi and Dewey's model of inquiry emphasizes the concept of equilibrium (cf. EK, sec. 1.5), but the specifically biological idea of (what we would today call) homeostasis is important only to Dewey, where it plays a role both in his analysis of life and in his epistemology (cognition being a homeostatic mechanism). States of organic equilibrium are disturbed, but then, through actions on the environment, equilibrium is restored. Thought on this picture is *among* the organic activities whose basic function is the restoration of equilibrium. Dewey's use of the latter notion reflects the close affinity and continuity that Dewey saw between life and mind generally. A state of organic disequilibrium is a state of *need*, and one might say that a state of *doubt* is nothing but the mental version of this same kind of structural pattern. As Godfrey-Smith (1998, p. 105) puts it:

> The disequilibria in living organisms induced by environmental events are like "proto-problems" for Dewey, and all living activities which act to regain the organic equilibrium are proto-solutions.

This does not tie in well with Levi's normativist – hence discontinuist – picture of the mind. Inquiry in its most general conception is a response to an "indeterminate situation" for Dewey, where the indeterminacy is crucially *recognized* by the agent, not *created*. We are talking about an *objective* indeterminacy and about inquiry as inducing an "objective change to the situation, transforming indeterminateness to determinateness" (Godfrey-Smith 1998, p. 113). That thought restores determinacy is an idea that might well, were it not for its naturalistic character, provide a foundation for Levi's idea that agents commit to the truth (and not just to maximal probability) after going through a situation of doubt (indeterminacy). The commitment to the truth would then be the moment where a new determinate base of thinking (and potentially

[9] See Godfrey-Smith 1998, ch. 4, for a discussion of Dewey's naturalism.

new problem solving) is restored. Truth having no foundation in Levi (cf. EK, ch. 1) and having no relational definition in terms of correspondence or ontology, either, may in fact amount to little more than determinacy itself: a stable state from the perspective of which a new indeterminacy can arise and be dealt with.

I like this way of putting things, because so many have wondered how, on Levi's legendary infallibilism and absolutist conception of knowledge (EK, sec. 1.6), where the agent is committed to rule out the falsehood of any logical consequences of what he or she fully believes as being no serious possibility, such a state could ever be rationally either risked or revised. If knowledge is no more than a point of determinacy in the sense just recommended, the pressure of this puzzle seems resolved – on naturalistic lines or on the grounds of the fundamentally practical needs of a natural creature. Few would doubt that an environmental problem posed to an animal – catching prey or preserving body temperature – never gets solved. Such problems are not held open for all times. Why should cognition be different, and humans keep an open mind on any issue at all times, reckoning, as a matter of principle, *any* odd possibility a serious one? Should we expect natural selection to engineer *such* a creature?

Incurring a commitment *can* be to decide over truth and falsity, then, and the uncertainty-inducing, *relational* question of whether, in a given state of determinacy, we really have hit on the truth, never arises. Settling the truth is what the state *consists in*. Comparing a given state of knowledge with reality is excluded almost by definition. Truth is no relational notion, and may be viewed either as a commitment to having dispositions to endorse certain consequences (those entailed by one's current state of full belief), or as a value that may be evoked when justifying a change of that state.

5. COMPETENCE AND COMMITMENT

Returning to human infants, we may say that they are naturally equipped with "systems of knowledge" about certain domains (language, physical space, social relations, mental states, etc.). Developmental biologists and psychologists speak of "innate expectations" about the structure of these domains, but the idea appears to be the same: Infants come to encounter objects, persons, languages, and so on, knowing a great deal about them.

Can we say as well that they encounter them *believing* a great deal about them? I see no motive for this particular move. The picture of the child forming particular beliefs or theories about some particular domain-specific problem seems a strange assimilation of a model of theory-formation in science to cognitive development in the child. Whatever the use of the apparatus of propositional attitudes in domains of adult problem solving, that apparatus should be dispensed with if it yields no further explanatory benefit in the description of the child's language faculty as simply a complex dynamical system in nature that undergoes certain state changes prior to maturing and stabilizing around the age of puberty. If language acquisition is basically a process of maturation that depends to a very minor extent on environmental feedback or cultural difference, and leads to success in all cases except severe pathology, there is no point applauding the child for its well-conducted "inquiry" once the feat is done. An *evaluation* of cognitive success according to principles of rationality appears misdirected, and the picture of the child as undergoing a "contract" when risking a new commitment in the light of its contingent values and goals even somewhat bizarre.[10] That does not mean that the terminology of belief and rational belief change yields no benefits in other respects, and I turn to these later on.

This picture involves a move from a commitment-performance distinction so crucial to Levi's picture to a *competence*-performance distinction. Performance is not as such subject to scientific study, as a vast set of cognitive faculties and background assumptions enters in even the simplest act of language use. Idealization and abstraction is the only way to study reality in its full richness. With this, it seems, we must agree, and idealization and abstraction will not bring us outside the confines of naturalistic inquiry. Evaluating performance against a background of *commitments*, by contrast, will. It is

[10] If our possessing grammatical competence is a biological feature of us having nothing to do with contingent values or rational performance in Levi's sense of these terms, it may be that our *logical* competence should be described in a similar way. Having a mind structured in whatever ways evolution structured it, we find us accepting certain inferences as compelling. On this view, it would be wrong to talk about our (on Levi's view, unrevisable; cf. CR, 17) "beliefs in the truths of logic." We have no such beliefs; we just have a mind with structures that turn out usable in various ways. Related to that, I also disagree with Levi (2003, p. 125) that the notion of (linguistic) meaning becomes redundant once the notion of belief and the apparatus of rational belief change is adopted. A naturalistic attempt to explain, on the basis of linguistic principles, why a particular sound structure corresponds to a particular meaning structure is untouched by work on the revision of our beliefs. Analyticity, too, as I have argued elsewhere (Hinzen 2004), should not be discussed in either the epistemological or the alethic mode. The immediate cognitive steps from *I painted Bettina blue* to *The surface of Bettina's body is blue* (not Bettina the person) or from *My Saab has a Ford T engine* to *My Saab runs by means of a Ford T engine* (rather than having a Ford T engine stored in its trunk, which is a possible meaning of *There is a Ford T engine in my Saab*) are not inferences guided by norms for how we rationally change our minds. They have explanations in terms of how evolution and system-internal constraints have built our shared human faculties of language or linguistic competence (for the explanation of the last example, see Hinzen 2003). To the extent that these explanations are correct, there is no need to assume that these analytic truths have anything to do with belief, cognitive evaluation, or how we change our minds.

with this methodological decision that I take issue, as it is this very step that deprives us of human rationality as a subject matter for naturalistic inquiry.

While having this sort of subject matter is a crucial aspect of both Cartesian and Chomskyan rationalism as I understand it, foundationalism of the sort that Levi's pragmatist axiomatically rejects is foreign to it. While some antifoundationalists today combat the "foundationalism" of the rationalists, there is a sense in which there wasn't ever any serious issue of foundationalism after the epistemological crisis of the late seventeenth and early eighteenth centuries.[11] But in a sense the entire Platonic rationalist conception of knowledge, as it transpires in the *Meno*, is antifoundationalist. Here knowledge comes for free, resting on no evidential foundation in experience. It is natural for humans qua humans, irrespective of training or formal education, to have certain systems of knowledge. In the light of some experiential triggering, knowledge "dawns in us like in a dream," as the *Meno* (85c, 9–11) puts it strikingly, not a particularly rational way of coming to knowledge, one should think. How one can misread rationalism as inaugurated in the *Meno* and as carried much further in Descartes' philosophy of science as a "foundationalism" seems obscure.

Rationalism as I understand it, then, is the thesis that having certain forms of knowledge is part of having the nature of a human being. Human rationality is part of human nature and is as much subject to empirical study as the latter is. A "rational psychology" is no more than a science of the mind based on the insight that human cognitive performance makes sense only in the light of certain posited systems of idealized competence, describable in terms of certain principles and rules that a human uses to analyze experience. Psychology is the study of knowledge, of which performance is only the kind of indirect evidence from which we have to take our clues. Whatever our use for a commitment-performance distinction, in the study of the mind the need for competence as an explanatory notion seems hard to escape.

6. Levi's Principle

"Levi's principle" (Hinzen 2000), according to which "deliberation crowds out prediction" (CR, iv, 31–2, 76–9), brings an interesting twist into the story told so far. Evidence for this principle is perfectly intuitive, on the assumption that principles of rational choice have their intended application in the agent's monitoring the rationality of his own decisions. Such efforts are vacuous if Levi's principle fails: Principles of rationality are meant to apply *within*

[11] I am indebted to correspondence with Noam Chomsky here.

deliberation and help an agent to reduce a set of options judged to be *feasible* to a subset of *admissible* ones. If within deliberation a choice is predicted, say, on the basis that the choice will be *rational* or that an *admissible* option will be chosen, the choice is vacuous. For the agent should thereby become certain that the option is chosen, and then no other choice is, as far as he or she is concerned, feasible.[12]

That said, note that in principle the opposite conclusion could be drawn from the preemption result just reached: Principles of rationality are not to be used in the agent's monitoring of his or her own decisions but in *other* agents' predictions and explanations of his or her actions, or the theorists' (Levi concedes as much, cf. CR, 26, 31, 35). That is, rather than concluding, from the use of rationality principles in deliberation, that they cannot serve for prediction, we could conclude, from their use in prediction, that they cannot be useful in deliberation. Indeed, if human actions are parts of the natural world, and this world is inherently probabilistic, why should human actions not be probabilistically predictable, whatever our first-person perspective on the matter? Of course, if principles of rationality are false as applied to human beings and thus not predictive, this is not an attractive line to take. But then, this is not an objection, for maybe one should *abandon* rationality theory as a theory of human behavior, rather than *reinterpreting* it in normative terms.[13]

On Levi's own view, if principles of rationality were applied to rational beings (which, to repeat, he thinks we are not), they *would* be predictive: A sequence of actions will follow deterministically from them. But they aren't, as a matter of empirical fact, so we must not assume we are rational beings, as long as we stick to our principles of rationality. We are committed not

[12] In fact, Levi's principle forbids that the agent assigns *any* probabilities on his potential acts, not just the extreme ones (CR, 32, 76–7). Rabinowicz (2002) argues that this strong version is based on a too tight connection between probabilities and betting rates. I have pointed out (Hinzen 2000) that the rather far-ranging consequences of Levi's principle in this strong form with respect to game theory and the explanatory power of its various solution concepts should be weighted against the clear use that these solution concepts appear to have, particularly when placed in an evolutionary (i.e., nonnormative) setting.

[13] Even given such a reinterpretation, it is not clear to me why a deliberating agent should not be allowed to assess the rationality of his or her acting in accordance with what principles of rationality predict: While deliberating, he or she might switch to another mode of thought, pausing to wonder how rationally he or she will likely choose today, given his or her past "record" of acting rationally. While these are indeed two different modes of thought, which would lend support to Levi's view, I can't see why it is rationality (rather than mere psychology) that prevents him or her from entertaining these two kinds of thought at the same time. This line of thought suggests that Rabinowicz's (2002, pp. 92–3) "screening off" suggestion for justifying Levi's principle won't do: It points to a merely psychological incapacity, whereas Levi intends his principle prescriptively.

to assume that we will act rationally (the "smugness assumption," CR, 31). Our necessary uncertainty of our rationality is the price we pay for having principles of rationality making, as is natural for such principles to make, predictions on the behavior of rational beings. But this feels like going a step too far: Being told one acts irrationally seems bad enough, but being told one must necessarily not assume to act rationally or to assume one's own confusion (else one could predict one's actions, hence could not choose) seems somewhat harder to swallow. Again, why save the theory of rationality from being a refuted theory by stipulating a new prescriptive principle forbidding its use in prediction?

We need a motive for Levi's principle other than the one that it, if not accepted, makes a given theory have wrong predictions. That principles of rationality make predictions for beings that are described by these principles is, after all, the first and natural thing to assume. Why on earth, if they apply to us, should they not be predictive? Well, we mentioned the standard reasons offered: limitations of computational capacity, for example. But one can accept these limitations without giving up on the descriptive status of principles of rationality and going for Levi's strictly normative model instead. In fact, the way that systems of competence interact with systems of performance that use these systems of competence provides a different account of the same phenomenon, without normativity entering. Thus, if we assume, with tradition, that our language faculty can be described as a combinatorial system consisting of primitives and rules – a grammar – then we find in this domain as well that there is "no end to human confusion and diversion" leading humans to produce ungrammatical sentences. Grammatical rules being recursive, for example, they allow us to construct ever longer and longer sentences. Memory limitations will eventually disallow us from processing them, but that will not as such be an objection to them being generated according to the rules and primitives of the grammar. The grammar system, we will conclude, interacts with other systems, leading our theory of grammatical competence to make wrong predictions in an infinity of cases. It will be explanatory all the same, for we will appeal to the rules of grammar as entering into the complexities of language use and as explaining an aspect of their full richness. No motivation for normativism arises, and the question looms large of why we should not go for a similar account in the case of the human "faculty of reasoning" (to use a probably very misleading term), rather than prevent predictive and explanatory uses by additional stipulations.

In a discussion motivating his principle, Levi (CR, 37) argues that norms of rationality are also not "blueprints for rational automata." That would exclude robots from the intended users of norms of rationality, but again I lack faith in the distinction made here, which seems to reflect a nonnaturalistic bias and dichotomy: Humans deliberate (use standards of rational health for self-criticism), robots don't. The point appears to be that rational automata function according to deterministic laws, so that laws of rationality applied to them would lead to predictions of rational choices. But then again, why not let principles of rational choice have their predictive uses, while arguing that, in the case of humans, the system described by these principles is embedded in a quite different cognitive architecture than it is in robots, leading to interaction effects naturally excluding predictive uses of the same kind?

Robots are different from humans, to be sure, but a naturalist will start by assuming these differences are empirical, not categorical: Unless we beg the question for antinaturalism, it is not that we are also, but not only, physical beings, while robots are only physical beings. Assuming naturalism, we cannot make the empirical differences depend on common-sense intuitions we have concerning ordinary words such as "choice" or "deliberation." For native speakers of English and probably all other languages "choice" as well as "deliberation" analytically contradict "deterministic prediction." Levi's principle takes its intuitive appeal from right here. But then, it is rational systems we wish to study – human beings – not the meanings or conceptual contents of common-sense terms. If the robot does not "deliberate," do we? Of course. But from another point of view, such concessions lead to no comfort. Deliberation is a process that can be functionally described. We are, for all we can tell, lumbering robots, even though we like to describe us differently, and no doubt are well advised to do so in our daily lives.

7. FINALE: RATIONAL BELIEFS IN A NATURALISTIC PERSPECTIVE

As Elster (1989) emphasizes in his discussion of rational choice, explanatory weakness is not at all, in general, a necessary consequence of predictive weakness. A mechanism may easily have explanatory uses, but not allow us to make predictions, precisely because it may interact with a number of other mechanisms that may disturb the functioning of the first. In other words, we may not be able to tell when one among a number of possible explanatory mechanisms is enacted. Taking this into account, it seems that although being interpretable as rational may well not be *constitutive* for being in a mental state, it may sometimes be *illuminating* to appeal, in describing the workings of the mind, to the operation of a *mechanism of rational choice*.

If that mechanism operates, the person's behavior will be optimally adapted to the given circumstances. In Elster's picture, that optimality is internally differentiated into three distinct optimizations: The person has optimized the time to spend on the collection of evidence; the person has formed the optimal beliefs given the evidence (rather than merely being led by wishful thinking, say); and the person has determined the optimal means for realizing his desires given his beliefs. The role of beliefs is essential here. Being in a state of belief means to look at the world in a way that one's perspective is not colored by features that the world would have only if it were as one desires it to be. I imagine this like a filtering process in which only those features remain that are not induced by my desires. It is clear that only specific productions of the language module could be usable by a belief-producing mechanism thus understood, not, for example, *I'll leap in through the window*, which is not my mind's construction of a fact, but is a projecting of a behavior of mine into future developments of the present state of affairs that are not yet factual.

Despite the fact that the rational choice mechanism points to a sheer luxury that nature affords (it's just fantastic to ever get so optimally adapted), there seem to be rather clear cases where the rational choice mechanism *has* a role in the explanation of behavior. In those cases, postulating rational beliefs as theoretical entities in a naturalistic account that play a role in the chemistry that produces an action does not seem far-fetched at all. There are also other types of behavior, which, although they are not rational in the full sense, involve beliefs in a naturalistic sense of the word. Take, for example, the mechanism in which the belief that I cannot get something that I desire causes me to stop desiring it (the so-called sour-grapes mechanism).

Still, the rational choice mechanism will not serve as a general theory of behavior. There are specifiable conditions where mechanisms are enacted that give no role to rational choice as an explanatory mechanism. As Albin (1998) shows, some decision problems have no rational solution for reasons of computational undecidability in the Gödelian sense, and also cannot be known not to have such solutions. In such cases, postulating rational expectations or the formation of rational beliefs seems vacuous. More down to earth, it may be a sheer *lack of opportunities* that dictates and explains a person's action. The economy of those actions will do without the mechanism of rational belief as well. For a last example, action explanations often appeal to the mechanism of weakness of the will, or the fact that although I desire something strongly, there are other desires that win over, simply because they develop a powerful psychic turbulence. In this explanatory mechanism beliefs again play no clear role.

A coherent view then may be this: The human mind is a toolbox of mechanisms of different types. In some, rational beliefs do play a role, in others they don't. If so, the category of belief will have to be investigated much more carefully with respect to the different ways in which the mind enacts mechanisms for choice.

We began with an analysis of Levi's normative conception of rationality, and expressed our desire, not only to view human behavior and its explanations as part of the natural world, but also to see principles describing our rationality as explaining why we act in the way we do. Levi's objections to the ordinary notion of belief's having a naturalistic correlate and entering the explanation of human action will not convince a methodological naturalist, not least in the light of recent findings in cognitive and developmental psychology. No clear case emerges for necessarily analyzing human rationality as intrinsically emanating from a set of commitments, as opposed to a number of cognitive and interacting competences that naturally belong to us as humans. If mental aspects belong to us as other organismic aspects do, a recommendation on what well-conducted inquiry or change of mind is should be paired with a study of the intrinsic structures that make up the mind we do not change.

REFERENCES

Albin, P. S. 1998. *Barriers and Bounds of Rationality*. Princeton University Press.

Baker, G., and K. J. Morris. 2002. *Descartes' Dualism*. London: Routledge.

Bjorklund, D. F., and A. D. Pellegrini. 2002. *The Origins of Human Nature: Evolutionary Developmental Biology*. Washington, D.C.

Chomsky, N. 2002. *On Nature and Language*, ed. A. Belletti and L. Rizzi. Cambridge: Cambridge University Press.

Conlisk, J. 1996. "Why Bounded Rationality?" *Journal of Economic Literature* 34, no. 2: 669–700.

Cosmides, L., and J. Tooby. 1992. *The Adapted Mind*. Oxford: Oxford University Press.

Dennett, D. 1995. *Darwin's Dangerous Idea*. New York: Penguin.

Elster, J. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Foley, D. K. 2003. "Rationality and Ideology in Economics." Manuscript, Department of Economics, New School University.

Godfrey-Smith, P. 1998. *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.

Harman, G. 1998. "Moral Philosophy and Linguistics." Ms., Princeton University, November 11.

Hinzen, W. 2000. Review of Levi, *The Covenant of Reason* (1997). *Erkenntnis* 52: 403–7.

Hinzen, W. 2003. "Truth's Fabric." *Mind and Language* 18, no. 2: 194–219.

Hinzen, W. 2004. "Der gegenwaertige Stand der Gebrauchstheorie der Bedeutung." In A. Fuhrmann and E. Olsson (eds.), *Pragmatisch Denken*, pp. 59–86. Heidelberg: Ontos-Verlag.

Hinzen, W. Forthcoming. *Mind Design and Minimal Syntax.* Oxford: Oxford University Press.

Hirschfeld, L., and S. Gelman, eds. 1994. *Mapping the Mind*. Cambridge: Cambridge University Press.

Hume, D. 1978. A *Treatise of Human Nature* (1739–40), ed. L. A. Selby-Bigge. Oxford: Clarendon.

Kuhlmeier, V., P. Bloom, and K. Wynn. 2004. "Do 5-month-old Infants See Humans as Material Objects?" *Cognition* 94: 95–103.

Lenneberg, E. 1967. *Biological Foundations of Language*. New York: Wiley.

Levi, I. 1980. *The Enterprise of Knowledge*. Cambridge, Mass.: MIT Press.

Levi, I. 1991. *The Fixation of Belief and Its Undoing*. Cambridge: Cambridge University Press.

Levi, I. 1997. *The Covenant of Reason*. Cambridge: Cambridge University Press.

Levi, I. 2003. "Seeking Truth." In W. Hinzen and H. Rott (eds.), *Belief and Meaning – Essays at the Interface*, pp. 119–38. Frankfurt: Hänsel-Hohenhausen.

Mehler, J., and E. Dupoux. 1994. *What Infants Know*. Blackwell: Oxford.

Pinker, S. 2003. *The Blank Slate: The Modern Denial of Human Nature*. New York: Penguin.

Rabinowicz, W. 2002. "Does Practical Deliberation Crowd Out Self-Prediction?" *Erkenntnis* 57, no. 2: 91–122.

Rorty, R. 1990. "The Priority of Democracy over Philosophy." In Alan R. Malachowski (ed.), *Reading Rorty*, pp. 279–302. Oxford: Blackwell.

Wertheimer, R. 1974. "Philosophy on Humanity." In R. L. Perkins (ed.), *Abortion: Pro and Con*. Cambridge, Mass.: Schenkman.

18

Psychoanalysis as Technology

Akeel Bilgrami

I

This chapter is about the relationship among three things: the concept of agency, the concept of mentality, and the practice and theory of psychoanalysis. Its large effort is to show what consequences follow for our understanding of what goes on in psychoanalysis, once we properly understand the nature of agency and the kind of mentality that makes agency possible. For some years, I have been exploring the relations between a normative account of the nature of agency and various issues in the philosophy of mind, most particularly the special nature of self-knowledge among all the knowledges we have. Since self-knowledge as a theme is so central to psychoanalysis, it is not surprising that that exploration should yield some consequences for how we might think of the basic notions by which we approach psychoanalysis. But it was not until I came first to see the importance of viewing mentality (in its intentional aspects) in terms of commitments, under the influence of my colleague, Isaac Levi, that I could quite see the interrelations between the notion of agency and the subject of this chapter. For some years he would hammer away at my efforts in these areas, insisting that my own instincts and ideas on the nature of agency and its effect on self-knowledge implied a particular view of intentional states as commitments. I resisted him for a long time until I realized what a flood of systematic illumination it indeed brought to my more piecemeal pursuits in these regions, once one embraced that very specific view.[1] Quite apart from this particular point in philosophy, I owe him an enormous intellectual debt for the constant encouragement, stimulus, criticism, and instruction that I have received from him as a colleague

[1] On the other side of these themes, the side of psychoanalysis, I have been much influenced by Garrett Deckel – not merely by her academic as well as intimate knowledge of the subject but by the sharp philosophical sense she brought to it. I have discussed the themes of this chapter with her for many years and we have written a version of it together, with an emphasis somewhat different from this one, which will be published in a different context.

at Columbia. His originality as a thinker and his purity as a philosopher, by which I mean his efforts to pursue his philosophical ideas without distraction from what is demanded by career and professional interest, is rare in our discipline, and the example he has set for us will become rarer over the years and is therefore the more to be respected and admired.

II

Some years ago the eminent psychoanalytic theorist Roy Schafer (in an influential book called *A New Language for Psychoanalysis*[2]) wrote to lament the tendency in psychoanalysis and its theorists to a discourse that leaves out too much the perspective of agency for a too passive understanding of human mentality and behavior, and he urged instead a turn or return to a more active voice. I want to salute that sentiment, but I am not at all sure that what I have to say by way of developing it in this chapter is what Schafer had in mind, or whether it will even please him and other psychoanalytical theorists very much. All the same I think it is strictly implied by the sentiment he expressed.

Let's begin with a conceit. Imagine, if you can, a subject that is completely passive. If it helps, think of Oblomov, the eponymous hero of Goncharov's novel, and then exaggerate him superlatively. What is meant by passive here? He is, likely, inert. But that is a relatively superficial thing. In fact, it is not a requirement that he be inert, since he could be blown around by gusts of wind and still be passive in the sense I have in mind. What he lacks is not movement or behavior but a certain point of view, what Kant called the point of view of agency. He thinks of himself exclusively as a product of his experience and its causes. He, therefore, thinks of his future just as he thinks of his past. He does not think that he can make any difference to it. This is the subject as object. The right way to describe him is to say that he lacks the "first-person point of view."

Now such an Oblomov as a conceit of imagination is, of course, hyperbole, and he may in fact not even be so much as imaginable. But it is the effort (in the name of science) to approximate in our theoretical descriptions what the conceit represents as an ideal, from which Schafer asks us to recoil. What lesson are we to learn, then, from his methodological advice? Nothing less than this: The more we approximate this ideal, the closer we are to negating the very idea of mind.

[2] Roy Schafer, *A New Language for Psychoanalysis* (New Haven: Yale University Press, 1976).

What relevance does this have for psychoanalysis? Its relevance can be sighted first in a paradox in the way we understand the effects of the disciplinary regimes of psychoanalytic and psychiatric disciplines. Ever since Spinoza and most explicitly since Freud, we are familiar with the idea that our agency, that is to say our autonomy and self-governance, is in fact enhanced by coming to understand the ways in which we are caused by various mental states, conscious and unconscious, to do the things we do. In other words the very way in which Oblomov thinks of himself – in the third person, as a product of such causes – when exercised in these disciplinary formats to unearth and bring to our own grasp the relevance of our own mental histories, is liberating. And the paradox is this. This third-person exercise that we perform on ourselves is the genuinely liberating thing that Spinoza and Freud claimed for it, only so long as it is not elevated into some sort of ideology about the nature of mind itself, only so long as it is not, as Schafer puts it, erected into a discourse or "language" of mentality. The conceit shows that decisively. Oblomov, after all, is the very antithesis of agency and autonomy, he is just the logical end, and the *ideological* by-product, of precisely such a third-person point of view. So we are presented with a deep but undiagnosed fact as to why the *very thing*, which, when writ small, *induces* agency, should, when writ large, *reduce* it.

Some of the systematic efforts of this chapter are intended to provide just that diagnosis.

A subject who, like our Oblomov, is completely passive, is a creature for whom thoughts, if he or she has them, are the sorts of things that happen to him or her. Being passive, he or she does not think them, since he or she does not *do* anything. His or her thoughts *assail* him or her. This idea, I believe, makes no sense; it literally makes nonsense of our notion of thought to say of a subject that *all* his or her thoughts are such that they merely assail him or her. In saying this I believe I am simply following the lead of Kant, for whom the perspective of agency was a necessary condition of both practical and theoretical judgment. The point can be put in a way that extends Kant's doctrine of transcendental idealism to hold not merely of objective experience of the world (of which Kant spoke explicitly and at length), but also of our experience of thought or intentionality. The idea is simple. On my view, this extension of Kant's doctrine rules out the following picture of thought: Thoughts are there anyway, and what the perspective of agency brings to them is a trigger or switch of activation. The phrase "there anyway" familiarly captures a highly *realist* notion of thought (a transcendental realist notion, as Kant would have said). According to Kant's transcendental *idealism*, thoughts are *not* there anyway *independent of the perspective of agency*, any more than the elements of objective experience of the world are there anyway, independent of our concept of, say, cause. Neither causality nor agency is a bit of "extra" brought to bear (respectively) on independently existing elements of objective experience and an independently existing realm of thoughts; rather, there is an integrated picture of the world and of thought, in which both causality and agency (respectively) are *constitutive conditions*, an integrated picture that Kant first profoundly brought to our attention.

Once it is so simply put, it may seem that I am pushing at open doors. It may seem obvious (to those who are not philosophers, anyway) that agency is, in this way, a necessary condition of thought. But it has not seemed so to most philosophers. That should not surprise us since, as Wittgenstein said of so many doctrines, they are so obviously false, that only a philosopher would have thought them up. What is it that philosophers think that contrives toward a denial of the obvious claim relating agency, necessarily, to thought? It's this: Most philosophers think that intentional states are holistically linked, propositionally specified, causal and dispositional states. Whole theories of the mind have been constructed with great sophistication and rigor that take for granted that intentional states are more or less systematic elements in a causal and functional picture of the mind. All of cognitive science is built on this assumption. But if intentional states were dispositions, it is hard to see why (or how) agency would be constitutive of them. Dispositions (at least as understood by this broadly functionalist picture) are presumably just the kinds of states which *are* "there anyway," and they need one or other triggering conditions to be activated. It is precisely because they are there anyway that one can expect that underlying them is some more categorical state – presumably physical – or at any rate what underlies them is a dispositional state of *a more fundamental* physical science. So just as the dispositional state of solubility that sugar has, for instance, has an underlying chemical basis, so also psychological dispositions have an underlying biochemical basis. And these physical bases are paradigmatically the kinds of states that are "there anyway," independent of our agency. And if intentional states are dispositions, there is no bar at all to thinking of Oblomov as having thoughts. He has all the dispositions that any one of us might have; they simply await activation. He is, therefore, full of potential behavior. He is fully minded.

To turn one's back on such a picture of mind would be to find a status for intentional states that is not dispositional at all. Of course, whatever that status is will have important links to one's causal and dispositional states. Intentionality is not something irrelevant to our motives, dispositions, and behavior. But it is not *itself* to be understood in causal, motivational, and dispositional terms. What, then, is this status apart?

III

Many philosophers have remarked and made much of the idea that thought or intentionality is a phenomenon that is "governed" by normative principles, and some – though by no means all – among them have said that it is irreducible to anything physical. But a surprising number of philosophers, even those who think of intentional states as irreducible, continue to think of them as dispositions. This position strikes me as deeply problematic. It is not clear at all how something dispositional can also itself be normative. Saul Kripke raises such a question sharply in his book on Wittgenstein[3] when he says we cannot have it both ways, and (though he spoils his own good point here by conflation of it with what he perceives to be a normativity in the *meaning* of words[4]) we might take his good point here as the point of departure for spelling out the distinctive status of intentional states. If intentional states cannot be normative by being dispositions that are – as the obscure phrase goes – "governed" by normativity, they had better *themselves be* normative states. A good word for them might be one that Isaac Levi himself first introduced, the word "commitments." Intentional states are commitments, and so I quickly need to say something about what these are, by saying why they stand apart from dispositions.

Commitments, to put the point as a bit of drama, are like promises to oneself. Our actions and other thoughts as well (but I restrict myself to actions in this chapter) can then be seen as having the role of either fulfilling those promises or failing to do so. Like promises, intentional states are fully normative states. That we have made a promise, that we have an intentional state, is a fact about us. But promises and intentional states are not themselves facts of nature. To say so would be to have missed their distinctive normative status, to have quite literally committed a naturalistic fallacy. They are therefore, simply not dispositions, in any interesting sense at all, dispositions as I said earlier, being paradigmatically, naturalistically understood phenomena. All beliefs and desires and other such states, when they are conceived as genuinely intentional states, are commitments to do or think various things. The desire to help the poor is a commitment to do various things: give money to charity, as it might be, or join a communist party. The belief that there is a table in front of me is a commitment to think various things, such as (to

[3] Saul Kripke, *Wittgenstein on Rules and Private Language* (Oxford: Blackwell, 1982).

[4] I have tried to show elsewhere why the normative nature of intentionality should not be carried over too uncritically into seeing the concept of meaning in normative terms, in particular, criticizing Kripke's own treatment of meaning in that book. See Akeel Bilgrami, *Belief and Meaning* (Oxford: Blackwell, 1992).

take just one) to think that there is something in front of me. A failure to do and think those things is perfectly compatible with having those corresponding commitments, with having that belief and that desire, since one's commitments don't cease to be commitments just because we do not live up to them. And for the same reason, even the failure to have the *disposition* to do or think those things is compatible with having those commitments. That is some sort of proof that commitments are not dispositional states. And it is only if intentional states are commitments, in this sense, that they can get to be normative in the requisite irreducible sense.

Oblomov, for all the dispositions he might have, in lacking agency, precisely lacks commitments, and that is why he lacks thought or intentionality. I'll return to the links between agency and commitments in a moment. I need first to say just a bit more about what these commitments are, with which I am equating intentional states.

For all I have said so far, it is not obvious that one can tell the difference between someone who has a certain intentional state and someone who lacks it. After all, if one does not have to *live up to* a commitment in order to have one, and if one does not even have to have the *disposition* to act so as to live up to it, then it is a real question of how one may distinguish the having of the commitment from lacking it. Simply placing the demand that we can verbalize the commitment with the words "I believe that *p*" or "I desire that *p*" will be insufficient, since the words may be phony, and the further demand that the words be sincere does not get us much beyond the initial question. That is to say, to ask, When does someone have a commitment? is not all that far from asking, When is someone's avowal of a commitment sincere? To put it as Polonius might have, "Actions speak louder than words," and it is precisely the actions, even potential actions, that go missing if there need be no disposition to act on a commitment in order to have one. So, evidently, some further demand must be placed on the idea of an intentional state before one can plausibly be said to think of it as a commitment.

Here is one. I put it down as a necessary condition of having a commitment. To have a commitment, one must be prepared to accept criticism if one fails to live up to it or if one lacks the disposition to live up to it, and one must try to do better by way of living up to it or cultivating the disposition to live up to it. There is a great deal more, in fact, volumes more, to be said about the nature of intentionality when viewed this way.[5]

[5] I have written in some detail about it in my book, *Self-Knowledge and Resentment* (Cambridge, Mass.: Harvard University Press, 2005). Levi has of course also written about it in detail and in depth in a number of books over his career. We have continuing marginal

A picture of thought or intentionality in which agency is constitutive requires, I have said, that the mind be normative, that it be peopled with commitments, and that Oblomov, as I have sketched him, lacks a mind in this sense; and so the question arises, What is it about agency (the property Schafer lauds) that requires this normative element?

IV

A lengthy excursus is necessary here about one of the oldest questions in philosophy: the nature of freedom and what precisely it has to do with value or norm. When I tried to portray a subject who completely lacks agency, I said that he lacks a certain point of view, a first-person point of view. I need to show briefly now why the notion of a first-person point of view – the notion that as agents we do not see ourselves from a third-person perspective where we are a product of causes and the trajectory of predictions, but see ourselves instead as thinkers and as actors, in short, as possessing freedom – has intrinsic links with precisely what makes our thoughts normative. How does the concept of value or norm provide the glue that joins thought (intentionality) to agency?

The problem of freedom and agency has traditionally arisen in the context of the question, How can there be free agency in the face of universal causality? Traditional answers to the question took two opposed forms, the first to deny universal causality, asserting that human actions in particular were free because they were not subject to it, and the second to assert that causality was exceptionless and assert also therefore that human freedom was illusory. These opposed views both shared a crucial assumption: that causality and agency were incompatible. A third position was then shaped that questioned the assumption, asserting that it was not *all* causality that threatened freedom, but only some particularly coercive causes. Causes of actions were not all coercive; in particular, causes such as our own beliefs and desires were not in the ordinary case coercive, and actions that they caused were therefore free.

disagreements about how to understand the nature of commitments. In particular, I am against describing what is implied by an intentional state (what is implied by a belief) as also being commitments, as he does. That is, a commitment (a belief, say) implies that one ought to believe various other things, but these latter, I claim, should not themselves be described as commitments. They have a status distinct from the belief that implies them, and one element of the distinctness is that they (unlike the belief that implies them) are not commitments. The reasons for this are complex and perhaps register a far-reaching disagreement between Levi and me, and I do not spell them out here. But despite disagreements such as this, I follow him in thinking that intentionality is to be thought of not in dispositional terms, but rather as commitments. These are, as he likes to say, "Trotskyite quarrels," once that main point is embraced.

Only actions that were caused by such causes, such as someone physically forcing someone or holding a gun to another's head, or internal causes such as raging urges to drink or smoke or steal, were unfree actions. So it was not causality per se that threatened agency, only *some* causes with the particular property of coerciveness or compulsiveness. This was a plausible position, but it was markedly incomplete. A question that needed to be addressed was: What makes a coercive cause coercive? Since all actions are caused, what singles out the freedom-threatening causes from the others? No satisfactory answer to this question surfaced until a brilliant essay by Strawson called *Freedom and Resentment*[6] pointed out that the coercive element in a cause could not be discovered by looking only at the cause. It could be discovered only by also looking at *our own reactions* to the actions that were caused by it. So what made an action free and the cause that produced it noncoercive was that our reactions to that action were reactions of blame, indignation, and resentment when it was harmful, and reactions of praise and admiration when it was worthy. By contrast, our reaction to even the most harmful of actions tended to be one of forgiveness or indifference if it was caused by a coercive cause. It was Strawson's point then that neither freedom nor the coerciveness that threatened it was a *metaphysical* property of actions and causes looked at from the outside, but rather it was a property of these things looked at only from the inside of our *evaluative* responses to certain actions and the sorts of causes from which they flowed. Agency therefore could be salvaged from causal determinism only if it was seen as a normative rather than a metaphysical notion. Without this element of value, we could not distinguish between coercive and uncoercive, free and unfree, and all actions would have to be seen as determined by causality, all freedom as illusory.

This strikes me as right, and deeply right. In fact, I think that Strawson does not take the normative element far enough by stopping where he does, by stopping at the *fact* of our reactive attitudes. Strawson says our actions are free because we just simply are creatures who have evaluative attitudes of resentment and admiration toward each other's actions. That is a *defining* feature of human beings, or as he puts it, inventing a term of art, of "persons." In saying this, he stops disappointingly short of the full radical implications of his own normativist insight. In saying this, he has no good answer to the following simple objection. The fact is that we have all sorts of reactive attitudes that perhaps we should not. I had a cat who used to urinate on my favorite volumes of poetry and I resented her for it. My wife resents her untuned piano. But we do not count cats and pianos as free.

6 P. F. Strawson, *Freedom and Resentment* (London: Methuen, 1974).

In the same vein, there are any number of psychiatry-driven ideologues who think that what this shows is that the converse lesson is generalizable to us, as human beings. The lesson is that we should not have reactive attitudes of blame and resentment and punishment toward human beings either, because all human action (just as much as the behavior of cats and pianos) is the product of coercive causes and dispositions. In short, we should cease to live such evaluative and judgmental lives since what we need is medical cure, not resentment and blame. What Strawson fails to address is such an ideological position. Against it, he complacently says that we just *are* evaluative creatures who can't help having these reactive attitudes. These attitudes are what make us what we are: "persons." What he should have said suggests a more radically normative position, one I recommend. And that is: It is not the case that we can't help having our reactive attitudes, but our own values are such that we think we ought not to give up our evaluative attitudes of resentment and indignation. We ought not to resent and blame cats and pianos, *because our further values* tell us that we should not. We should continue to distinguish between some human actions and not others as blameworthy because our further values justify doing so.

This is a more radically normative or evaluative position, because it does not see our being free agents and evaluative creatures as a resting point, as a mere primitive *fact* about us, but internally justifies our being evaluative creatures by appealing to further values we have. It is precisely the *failure* to find any values that justify our reactive attitudes that leads to the sort of alienated despair that Eliot, for example, was portraying in *The Wasteland* or that is found in our caricatured idea of decadence (of Ancient Rome just before the "barbarians" overran it), where we gratify only our sensations, but feel that nothing matters enough, no value promotes any judgment of praise or blame. The surrender of value and of judgment being portrayed in *The Wasteland* and supposedly in Ancient Rome is precisely intended as portraying a surrender of agency itself. My students call it a sort of pervasive "vegging out." More ceremoniously, we might call it a sort of "rational suicide" and formulate a conceit such as Oblomov.

This, then, is where Strawson's initial line of inquiry lands us, and indeed where Schafer's methodological advice lands us. The agency Schafer demands lands us as far as the idea that we are creatures with commitments as well as dispositions, and we have commitments not because we cannot help having them, as Strawson says, but because our further commitments promote them. Commitments are supportable only by further commitments, and *that is why they are irreducible to any facts or dispositions of nature*. There is nothing to sustain our being creatures of commitments, nothing to stop our becoming decadents and "wastelanders," nothing to stop ourselves from committing rational suicide, nothing to stop us from becoming Oblomovs but further commitments of ours.

I can now say something tentative about the paradox I had posed at the outset. A third-person perspective by which we understand the underlying causes of our actions can enhance our agency rather than destroy it, so long as we have *values* that reject the elevation of this perspective to an ideology, that is to say so long as we have values that reject the comprehensive psychiatry-driven medical attitude toward ourselves, whereby we surrender our reactive capacities altogether.

The idea that it is *rational* (not biological) suicide we commit if our commitments cannot justify our own agency and evaluative attitudes suggests that *reason* is the sort of thing that is concerned with commitments. And that is my starting point for a discussion of psychoanalysis. I am sorry that it has taken so long to set the stage. Actually, one final prop still remains.

V

The large frameworking thesis of this chapter is that psychoanalysis, as a disciplinary regime, is not concerned with reasoning directly or with values and commitments, though it indispensably presupposes them in the background, with the foreground itself being nothing but a sort of rarefied technology.

I have said that the domain of reasoning is commitments, but I have not said what exactly I mean by reasoning and why it is restricted to this domain.

The codifications of reason are famous and familiar. There are the codifications of deductive and inductive rationality, both of which address relations between cognitive states, such as beliefs. There are the codifications of decision theory that highly refine the basic idea behind Aristotle's practical syllogism, and they address the relations among beliefs, desires, and choices. (There ought to be some codification, however primitive, that systematically links the rational relations that hold among desires themselves or among values. These relations would have to be relations of coherence analogous in some way – though which exact way needs to be thought through – to the coherence relations that hold between beliefs. That is a field not much worked on, but desperately in need of work.)

All these codifications are about normative relations that hold between mental states, thereby making those states the irreducibly normative states they are. When one *reasons* with someone, with a view, say, to changing their mind, one tells them that they have some commitment (that one thinks they *ought* to shed) because it does not follow deductively or inductively or does not cohere with other commitments of theirs. Or one tells them that they do not have a commitment (that they *ought* to acquire) because it follows deductively, inductively, or coherentistically from other commitments of theirs. Or one tells them that they *ought* to choose to act in a certain way that they have not, or ought not to choose something that they have, because that is what is sanctioned by the codified norms of coherence and decision theory to be in accord with their commitments.

None of these questions of rationality apply to Oblomov, and none of them could. Why? Because all these intentional states that the norms of rationality address, states such as beliefs and desires that stand as relata in normative relations of coherence, consistency, confirmation by induction, and so on, are states on which we cannot have an exclusively third-person angle, and so we cannot merely be passively assailed by them, at best observing them when that happens, but taking no more active stance toward them. In other words, they cannot be dispositions, which Oblomov has in abundance; they have to be commitments, of which he has none. Its not that we *cannot* have a third-person angle on commitments. I can say or think to myself, "That is a commitment I have" as a bit of observation. But of course observation is not active, in the sense that Schafer demands. It is passive, since to observe oneself is to treat oneself as another. But that is not the only angle that we have on our intentional states. What makes them intentional states proper is that they are the sorts of things on which we can also have a first-person angle, precisely what Oblomov lacks, an angle of observation not as a bystander, but rather as an endorser, one who takes it up as a commitment, and *therefore* can be the justifiable object of someone's (including one's own) reactive attitudes, thereby making one an agent.

This, then, is how *value, freedom*, and *the first-person perspective* (a perspective that makes self-knowledge different from all other knowledge, such as observational knowledge of the world) all reduce to a single and highly integrated package of elements, all of which Oblomov lacks. Lacking one, he lacks all three since these are not separable elements. What I am taking pains to say, then, is that Schafer's methodological proposal, though he does not present it that way, is a highly omnibus one: It integrates with other elements that he does not explicitly draw out himself. On the face of it, his proposal seems to be about agency. But we have found underneath that agency is equally about value and about the nature of the unique angle that only we can have on ourselves. Let's put the point another way. We might have thought that there are three distinctions of interest: first, the distinction between agency or freedom as opposed to passivity or determinism; second, the distinction between norm or value and natural fact; and third, the distinction between our own angles on ourselves (the first-person angle) as opposed to our angle on others and the world (the third-person, or observer's, angle). But if I am right, these are all at bottom the same distinction.[7] There is no freedom and agency if all we have are dispositions and no evaluative states, and there is no need for a first-person perspective if we are passive creatures with only dispositions and no normative states. The implications of Schafer's remark that we need to acknowledge agency, therefore, brings with it an acknowledgment of *much else* besides.

The idea of rationality issues directly from saying that our intentional states are commitments because it is rational to be in accord with your commitments and irrational to fail to be in accord with them. In this picture of rationality, what is rational does not come from any standard *outside* our own commitments. We may be committed to certain things quite rationally so long as they cohere with other commitments of ours, even if they do not square with the trends and tenets of culture and morality. Of course, what we are committed to could be much influenced by those things; that is not ruled out as an empirical fact about us, and is no doubt often true of us, but there is nothing about rationality so conceived that necessarily issues from influences outside. Rationality itself is internal to the commitments of the individual, even if many commitments he or she embraces are influenced by external standards.

VI

The relevance of this internalist picture of rationality to psychoanalysis is that a person's mental well-being can now be seen as coinciding with his or her rationality, *so defined*. Since the demands of rationality are not demands that can be external and therefore alien to him or her, they merely tell him or her what he or she believes, values, and wants and what their implications are for his or her actions and his or her other beliefs, values, and wants. So also *failures* of mental health coincide with failures of accord with one's commitments, in one's actions, in one's dispositions, and in one's commitments themselves when they lack coherence with each other.

I have mentioned two broad kinds of failures: when our actions and dispositions do not accord with our commitments, and when our commitments do not cohere with our other commitments. The latter form of irrationality is fundamentally different from the former. Why? Because only reasoning will

[7] In this I am adding to the joining of intentionality with norm (in the notion of commitment as we find it in Levi) a joining of both of these with agency.

alter the situation, since both sides of the conflict involve normative states. The former are a different matter altogether, since they are a form of irrationality in which things that are not essentially normative, things such as dispositions and the motives that cause our actions, are in conflict with our commitments. Here reason has only a limited function. It can point out to us that conflict exists, and it can point out to us that we ought to get rid of it, by changing our dispositions and motives. But that is not sufficient to bring about such a change. The rest comes from control, and control sometimes is hard to achieve without some form of what I am, provocatively, calling technology.

Why use this nomenclature? When one is dealing with changes of normative states, only deliberation about them is to the point. But the kind of irrationality in which control is necessary has to deal not with normative states but with purely causal states, dispositions, and motives, to which deliberation cannot possibly make any difference, after a point. What is needed is to bring dispositions into line with what our commitments sanction, and bringing them into line means either curbing or getting rid of the dispositions that conflict with our commitments, or acquiring dispositions that we lack but which our commitments require (acquiring them, say, by cultivation of certain *habits* of behavior). All this involves not reasoning but drill of one kind or another. That the drill is psychological and not physical should not mislead us into thinking that it is not a technology. The idea that if it is not physical (if it is not manipulation, say, by hypnosis or by medication), it must be deliberative is an impoverished conception of the options. If this is right, the refinements that Freud visited on his own methods from the early phase of post-hypnotic suggestion to the fully psychoanalytic methodology is a sophistication and development *within* a technology, not a shift in direction from technology or manipulation (hypnosis) to deliberation (psychoanalysis).

Nothing in the idea of a technique requires that it must be physical. Perhaps in the fullness of time and knowledge, eventually only medication will be necessary to bring our dispositions into line with our commitments. But until we have that full knowledge, we have to exercise forms of technique that are psychological rather than physical. The important point is that we should not get confused into thinking therefore that just because we lack the full physical knowledge, that until we acquire it, what we are doing is something other than technology. The distinction between the physical and psychological does not coincide with the distinction between that to which technology is relevant and that to which deliberation and reason are relevant. This is because, as I have said, the psychological is a domain that is deeply divided by the dual aspect of the intentional and the dispositional. Because the latter aspect requires neither norm nor the first-person deliberative angle, both of which constitute agency, it is an aspect that is quite as susceptible to technology as the realm of the physical. It is just that the disciplinary regimes and techniques are bound to be very different and perhaps far more interesting and subtle than anything in the physical realm.

Some of the interest and subtlety comes from the fact that the realm of dispositions and motives is a realm that, while we are in a position of epistemic weakness, that is to say, while we lack the full reductive or at any rate integrative knowledge of its categorical, biochemical basis, leaves us no other option but to use propositions to specify what they are. Thus it is that we describe mental dispositions in contentful terms just as we do normative states such as commitments. That is the source, or one source, of the widespread conflation of intentional states with dispositions, which Kripke was descrying. Given the inevitable specification of dispositions in contentful linguistic terms, while we are epistemically weak, any efforts we may bring to the task of aligning our dispositions with our commitments will involve techniques that are radically different from material technologies. The technologies after all are addressing not the balance of chemicals, nor the configurations of cells and neurons, but propositionally specified phenomena, by which we don't mean just linguistically described phenomena, but phenomena that take linguistic objects such as propositions as their contents.

Dispositions that get such a description are thus rightly described as objects of interpretation and as having meaning. But, if I am right, the interpretation of dispositions by assigning meaning and propositional contents to them is, in one crucial sense of the term, "unprincipled" – since it is something that we do out of a position of epistemic weakness. It is an *instrument* by which we can come to an understanding of them in our efforts to bring them in line with our commitments. In short the interpretation of dispositions, the assigning of propositional content to them, is an *instrumental* matter, not a matter of principled description of a phenomenon that is stably and ultimately *irreducibly* propositional. It lacks that full prestige that is reserved for what is irreducibly normative. This means that the domain to which psychoanalysis gives its most assiduous and creative attention – the unconscious – involves interpretation that is purely instrumental, for unconscious mental states cannot possibly have the normative property of commitments as I have defined them. Why is that?

The defining property of commitments is not that we are disposed to act on our commitments. If it were, our unconscious mental states could well be commitments, since there is no gainsaying the fact that many of our dispositions are unconscious. The defining property of commitments is that they and the acts they might lead to are the sorts of things to which we have justifiable reactive attitudes of praise and blame and indignation and resentment, and that we ourselves are prepared to accept criticism for failing to live up to them, when we do, and to try to do better. But none of these things is appropriate to unconscious mental states, since it makes no sense to say that I am prepared to accept criticism for a state of which I am unconscious and would deny as possessing. These states are fully dispositional but altogether lack the defining normative element that makes for genuine intentionality. Even where relative coherence exists among our unconscious states, their intentionality is only a form of *mimicry* of the real thing.

A possible objection to this might come from the phenomenon of unconscious guilt. It might seem that if we have a first-order unconscious state (say, a hostile and poor opinion of a parent) and feel some unconscious guilt about having it, then we have the sort of self-criticism and self-reactive attitude at the level of the unconscious that should allow that first-order unconscious state (the belief about the parent) to meet my defining condition of a commitment. But the fact is that the guilt itself is not something we could see as being rightly or wrongly held, if we were not aware of having it. No praise or blame for having it or for actions that flow from it would be justifiable. The normative element at the unconscious level is at best superficial, at worse, and more correctly, fake. In that sense even guilt *that we are aware of* but that we have not yet endorsed into a commitment (a genuinely normative state) is just a further disposition, and about it we can raise a perfectly good normative question: Ought one to feel it? – the only question that is genuinely normative. Guilt itself does *not* pose that question about the phenomenon one is feeling guilty about unless the guilt has been endorsed into a commitment. It is only the latter therefore, which is a genuinely normative phenomenon.

Almost from the moment that Freud formulated his systematic thoughts on the unconscious, philosophers and others have wrestled with the question of whether intentional states can be unconscious, and many have doubted that they can be so. But none of those who have doubted it has given a completely convincing reason for doing so, because every one of them (whom I have read, anyway) has also assumed that intentional states are dispositions, in some sophisticated sense. On that assumption, nothing whatever could prevent the unconscious from containing intentionality. Nothing we could add to the idea of a disposition by way of sophisticating it could make a principled requirement that one *must* have self-awareness of it, nothing at any rate that does not sophisticate it so much that it is no longer recognizable as a disposition. Those who are intuitively skeptical of the idea of unconscious intentional states, I think, will find *no principle* to ground their skepticism unless they come around to the criterion for intentionality that I have proposed, that states possessing intentionality are themselves commitments defined as the sort of thing that we are prepared to accept criticism for, and to try and do better by when we do not live up to them. One could of course complain about me that I am ruling out too much that is intentional by insisting on this version of the normative criterion for it. But the onus is on those who make this complaint to show how any notion of the intentional that is more accommodating than mine does not fall afoul of Kripke's (and indeed Levi's) demand that we distinguish the intentional from the merely dispositional.

It is a highly revealing fact, however, that such a high level of mimicry can exist at all at the unconscious level. And one thing it reveals is that a technology that we bring to bear on unconscious mental states (states that I have been saying are necessarily dispositional and only intentional manqué) must work on these dispositions holistically, and speak to the patterns of inferential links that seem exemplified by these dispositions. A great deal of the techniques of psychoanalysis are geared to do just that, both in the bringing to the surface of awareness unconscious dispositions and in the aligning of these dispositions (once they are brought to the surface) with the commitments, when they conflict with them.

VII

A very crude framework is now emerging within which we may see the relations between the subject of psychoanalysis and the larger setting of mentality in its intentional aspects. Something roughly with the following elements.

- Our conscious mentality consists partly of intentional states, properly so-called, and not merely instrumentally attributed. These states are inherently normative and not merely dispositions that seem to exemplify the patterns of inference. They are commitments either to actions of one or other sort or to other mental states, which are implied by the norms of one or other codified forms of reasoning.
- There are other mental states, which are not intentional, but dispositional; they are like tendencies in physical nature, such as elasticity or solubility, but while we are in a position of epistemic weakness regarding their biochemical basis, they are necessarily of a far greater subtlety and complexity than straightforwardly physical dispositions, and they need contentful description mimicking the propositional specification of intentional states such as beliefs and desires, properly so-called.

- There can be deep conflict between our commitments and our dispositions, which can give rise to anxieties and neuroses of a wide variety.
- Some of these conflicts are between our commitments and *unconscious* dispositions, and the neuroses *they* give rise to are typically of the sort, which decades of sophisticated theory, building on the pioneering work of Freud, have studied. All neuroses fit this general framework. That is to say, ex hypothesi, the idea of a neurosis is a special case of the sort of disequilibrium that comes from a clash of dispositions with "commitments," in the broadly defined sense I have given that term.

All this suggests, roughly speaking again, that the analytical method has three central *conceptual* moments (though, of course, the actual *temporal* chronology is bound to be much untidier, involving much back and forth among these moments).

First, the discovery by the patient of some unconscious mental state (or set of inferentially linked states), which are necessarily dispositional. The path to this discovery is via a technology, a set of techniques, which have been much studied.

At this point, a *second* conceptual moment sets in, which is normative and not technological, and it is where deliberation takes place that requires the analysand to consider first whether the discovered mental state is in conflict with his or her commitments and, if it is, to see whether he or she wishes to bring the conflicted disposition in alignment with his or her commitments by changing the commitment or by changing the disposition. All this falls in the intentional realm, in the full and genuine (that is to say, noninstrumental) sense of that term.

Wittgenstein once said in his critical remarks on Freud that, ultimately, *assent* on the part of the analysand is the only criterion for the existence of unconscious states. Presumably, he meant potential assent, and if so, the point is not all that controversial. The deeper point is that he leaves it completely unclear whether by assent he meant what I have in mind by the first conceptual moment or the second. For those are two quite different things: to assent to whether one has a disposition that has hitherto been unconscious (the first moment) and to assent in the sense of endorse that disposition, thereby making it a commitment (the second moment).

In this second stage, one could endorse the disposition that one has discovered oneself to have or one could reject it. If one endorses it one has made it a commitment. That *is* what it is to endorse. The only thing left to reach an equilibrium is to make sure that this newly acquired commitment does not conflict with other commitments, and that task falls within the normative realm of reasoning. However, if one does not endorse the disposition but rather rejects it, there may well be more to be done, and what more there is to be done has nothing to do with reasoning, and further technology is required.

That is the *third* central moment of psychoanalysis: the path from self-knowledge of one's hitherto unconscious states to the actual reaching of a situation of equilibrium, the removing of conflict by getting rid of the conflicting disposition that one's commitments have rejected. The technologies involved in this third stage are also much studied by the theory of psychoanalysis.

So technology is relevant to two key moments, the first and the third: the first during which unconscious states, necessarily dispositional states, are uncovered, and the third during which, if certain discovered dispositions are rejected in the second normative conceptual moment, then they need to be actually discarded from the psychological economy by the techniques of this third stage.

To take just one salient example, transference (not the everyday phenomenon, but the clinical one) provides just such a technology relevant to the first and third conceptual moments. It is a certain widely studied reliving, which can, in the first stage, bring to the surface of a patient's mind a number of mental states (necessarily dispositional) hitherto unconscious, and moreover, once so available to the agent, Freud describes the *"working through"* again via transference as something necessary for the patient to undergo in order to be relieved of the disposition. That is the third stage. Thus during transference, the reliving brings to the surface long repressed states of mind, and the continuing of the transference past this academic knowledge of these hitherto repressed states helps the patient to see the gaps between the past and the reliving of it, so as to "work through," in Freud's phrase, those mental states, and thereby to ease them out of one's psychological economy.

Here are Freud's own words for the manifestly technological elements in the third conceptual moment I am stressing. They are from "Recollecting, Repetition and Working-Through."

> The first step toward overcoming the resistance is made by the analyst's discovering the resistance, which is never recognized by the patient, and acquainting him with it. Now it seems that beginners in analytical practice are inclined to look upon this as the end of the work. I have often been asked to advise upon cases in which the physician complained that the resistance had been made aware to the patient and all the

same no change had set in. . . . The gloomy foreboding has always proved mistaken. The treatment as a rule was progressing quite satisfactorily. Only the analyst had forgotten that naming the resistance could not result in its immediate suspension. One must allow the patient time to get to know its resistance of which he is ignorant, to "*work through*" *it*, in order to overcome it . . . only when *it has come to its height* can one, with the patient's cooperation, discover the repressed instinctual trends which are feeding the resistance, and only by living through them will the patient be convinced by their existence and power. This "working through" of the resistance may in practice amount to an arduous task for the patient and a trial of patience for the analyst. Nevertheless it is the part of the work that effects the greatest changes in the patient and which distinguishes analytic treatment from every kind of suggestive treatment.[8]

In early writings, Freud often talked of "catharsis." That this earlier terminology denoted a kind of technology (by contrast with deliberation) is perhaps obvious. But the later idea of "working through," as this passage shows, is not in essence different from it. In both, one is describing techniques for controlling dispositions. The entire vocabulary here and in several other passages is one of force (of unwelcome dispositions) and counterforce (of "working through" via transference). Technology against natural forces.

In his autobiographical study of 1925,[9] Freud explicitly uses both terms ("catharsis" and "working through") in the context of discussing transference, and then in a more general context explicitly links the two ideas when discussing the passage from hypnosis to analysis proper, under the influence of Bernheim. In neither the idea of catharsis nor that of "working through" is there any central (or even peripheral) place given to *reasoning* with a patient to curb or remove one or other mental state or attitude. It's rather a matter of letting the disposition "*come to its height*" and thereby having it eased out of you; and Freud is explicit in saying that transference is one absolutely central *technique* in this task.

Many more examples could and should be given to bring home decisively this point about technology, but I cannot do so here. In fact I have no doubt that I have been crude here in the hope of making the basic elements in the framework come to light, and come to light in the space of a short chapter. A great deal needs to be said by way of qualification, further examples, and other detail. I have even less doubt, however, that those details would not impeach the general claims of the framework.

[8] "Repetition, Recollection and Working Through" (1914), in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 12 (London: Hogarth, 1953–74), pp. 147–56.

[9] "An Autobiographical Study" (1925), in the *Standard Edition*, vol. 20, pp. 7–74.

VIII

If the last stage is pure technology, that means that the normative element is over once one has deliberated as to whether an uncovered mental state – some feeling of hostility, say – is to be endorsed as consistent and coherent with one's commitments or rejected as conflicting with them. Perhaps no one here needs convincing that the last stage is a technology. It may simply be obvious just from reading the various descriptions of (or undergoing) the dynamics of the analytical process. My point, however, is this. We do not have a theoretical *right* to this seemingly obvious conviction until we grasp the full contrast with the notion of agency that Schafer urges on us and all that it implies about the normativity of mind.

Psychoanalysis is pervasively silent on the matter of the normative element, focusing all its theoretical and strategic energies on the first and third moments, where the tools of uncovering states of mind and then working through them are most relevant. But the second stage is indispensable. If there were no second stage, the whole process could be performed by an Oblomov, for the first and third stages require no more than an observational or third-person perspective on oneself. Agency issues in the second stage, and that is vitally necessary for what allows the first and third stage to be the sort of liberating stages that Spinoza and Freud had claimed that they would be. It is the presence of the second stage that prevents the psychoanalytical method from deteriorating into the ideological picture of the mind, which creates the paradox I mentioned at the beginning of the chapter.

In short, the normative element is an unerasable backdrop, both to the very idea of what counts as a neurosis in the first place and to the task of achieving or approximating the goal of equilibrium that is supposed to ease the neurosis. Why is it crucial to defining neurosis? Because *accord of one's dispositions with one's commitments* is the standard of health, departure from which is a necessary (though by no means sufficient) condition for neurosis to even so much as arise. Without a standard of health, there could be no lack of mental health, no notion of neurosis. That specific kind of accord alone provides our idea of health, which is relevant to psychoanalysis. Nothing else can provide the relevant standard. Mere accord between dispositions could not be the standard of health, since consistent and coherent dispositions may still conflict with one's commitments. For that matter, mere accord among commitments could not be either the *complete* standard of health or the *relevant* one. It is irrelevant because even if our commitments are conflicted, that normatively unsatisfactory state is something against which psychoanalysis is helpless. It being entirely in the normative realm, only reasoning could resolve it.

But coherence among commitments is in any case not a complete standard of health, simply because one could have perfectly coherent commitments and yet have dispositions that conflicted with them. It is the alignment of *disposition with commitment* that provides that further standard of health, departure from which may give rise to the kind of neuroses to which the techniques of psychoanalysis are relevant. And the point is that psychoanalysis would have no subject matter if there was not this background of a standard of health provided by the normative ideal of the equilibrium or alignment itself.

Freud himself, perhaps because of his avowed distaste for philosophy, did not say much about this element of norm or value that lay behind the entire point of his repertory of structural claims, empirical hypotheses, and techniques for discovery and cure. They were unspoken assumptions. Part of the reason for the silence may well have come from two different conceptual theses he seems to have held about the nature of value, which distracted him from seeing the role for value that I am stressing in this chapter.

First, the notion of a norm he often wrote about was an external one, norm as morality, as issuing from conventional, social, and public demands. But that is not the notion of norm that surfaces in the framework I am placing on his ideas at all. That is not the notion of norm that is built into the idea of an intentional state, the notion of norm that comes from the idea of a commitment. Here it is entirely *internal* to the psychological economy of an agent, even if external influences may shape what one's commitments are, as they are bound to. Once the influence is in place, the point is that the notion of norm comes in with the idea of *accord with* one's *own* commitments. The normative idea of that accord or failure of accord is not at all the normative idea of an external moral value.

Second, Freud had a reductive view of norms as a result of thinking of it as having its source in an id-restricting external morality. He thought of norm (or conscience, as he often called it in *Future of an Illusion*[10]) as a second-order disposition or drive whose function is to curb one's first-order drives, which are highly destabilizing. This is a thoroughly disappointing, naturalistic view of normativity, which reveals the very point where Freud's scientific ambitions became outsize. It is not the point on which those who think he should be a hermeneuticist or a narrativist rather than a scientist focus, but the point where he simply thought that norms were themselves natural facts, second-order dispositions. Those who hanker for narratives and hermeneutics are saying something banal and perfectly compatible with Freud the scientist. Of course we *interpret* our unconscious states of mind and *tell a story* about

[10] "The Future of an Illusion" (1927), in the *Standard Edition*, vol. 21.

what underlies our behavior. How could this possibly conflict with anything scientific? Dispositions, so long as we are epistemically positioned so as to know nothing or nothing much about their detailed biochemical basis, leave us no choice but to see them as objects of interpretation, to which we assign meaningful propositions with inferential links to one another. That does not mean that dispositions are irreducible to natural facts. What is the argument that they are not? Mere dispositions do not even yield agency as the Oblomov conceit makes clear, so why should it provide any principled obstacle to scientific treatment? The fact that we have to exercise our hermeneutical abilities to grasp what these dispositions are until we gain more scientific knowledge does not support their irreducibility, in principle. It's only commitments that are the genuinely normative elements that come with agency and the first-person perspective, which is unyielding to reduction, and it is this point that escaped Freud. Keen to put in their place the demands of external morality, about which the author of this chapter takes no stand, Freud confusedly then also aspired to reduce the very idea of value to second-order drives. That is misplaced science.

Notice that this charge of misplaced science here has nothing to do with the shrill attacks in recent years on his scientific claims. I have been silent on the question of the verifiability and falsifiablity of particular empirical hypotheses in Freud. My chapter is concerned not with those specific claims in Freud, but rather with the larger framework in which his most general concepts and techniques are to be understood. In that framework, I am saying, norm and disposition both have an indispensable place: norm in the background, as the source and ground of the standard of health: disposition in the foreground, where the central energies of the theory and practice of psychoanalysis are focused. That Freud himself had a reductive attitude toward what I find to be an irreducibly present backdrop in the framework was, I claim, the only flaw of a headstrong scientific ambition. That the rest should be pure technology seems to me not merely right, but right only *because* the background is irreducibly normative.

Schafer had proposed that the concept and the conceptual vocabulary of agency should govern our understanding of psychoanalysis. I have tried to redeem that proposal in a specific way, showing how if we understand agency properly in all its implications, that introduces a normative, first-person element as standing behind a way of understanding the basic elements of neurosis and the various third-person technological, and nonnormative forms of dealing with it. This reconciles what it is that some theorists find both attractively scientific about psychoanalysis with what other theorists insist should be irreducibly humanistic about the nature of mentality. Once we see the relations between the normative and the technological in this unified framework, many of these old disputes in psychoanalytical theory between the scientific, on the one hand, and the humanistic or narrativist or hermeneutical, on the other, should be seen as unnecessarily adversarial.

I don't doubt that the irreducibly normative, agential, first-person element I think of as an indispensable part of the framework will raise the Crews and the Grunbaums to new heights of denunciation. They will no doubt find it to be a distastefully unscientific element. But that denunciation would no longer be targeting purportedly empirical hypotheses that turn out on inspection to be unempirical because unfalsifiable. That criticism is now wholly beside the point since the normative element relevant to psychoanalysis, unlike the dispositional element relevant to it, is not pretending to be scientifically or causally tractable. It was precisely my point that it is just as wrong (and wrong in mirror-image of one another) to say that the normative element in the background falls within the naturalistic domain as it is to say that the dispositional element falls in anything but the technological domain. The normative domain therefore frankly announces itself as standing in a principled way *outside* the purview of scientific ambitions. Grunbaum and Crews – for all I say in this chapter – may be right in some of the things they target for their criticism. But there is nothing in what I do say here that provides them with a target for their criticism. After all, as philosophers have pointed out before, it cannot be unscientific to insist that not all themes are scientific themes. It is only unscientific to give unscientific responses to *science's* themes.[11]

[11] As, for instance, is done by creationism.

19

Levi on Money Pumps and Diachronic Dutch Books

Wlodek Rabinowicz

It is with a great pleasure but also with some misgivings that I contribute to this volume. The pleasure comes from my feelings of friendship and gratitude toward Isaac Levi. We have known each other for a long time now. As I very well recall, it all started way back in the 1970s with his letter commenting on an article of mine dealing with his seminal *Gambling with Truth*. As a young and shy graduate student in Uppsala, I felt both overwhelmed and overjoyed by this great man's attention and encouragement. Suddenly, the distance between the faraway Columbia and my own university shrank to the manageable size of a philosophical argument. Thanks to Isaac, I realized, for the first time, that it was – perhaps – within my reach to join a larger community of minds that spanned the globe.

The pleasure is mixed with misgivings. Over the years my friendship and affection for Isaac deepened and matured, but philosophically we often found ourselves on opposite sides. He was highly critical of causal decision theory and I was one of its enthusiastic defenders; he was (and still is) a powerful advocate of the thesis that practical deliberation crowds out self-prediction, while I have been one of the doubters. Examples could be multiplied. In this chapter, as it happens, I want to examine another such bone of contention, more precisely the status of diachronic pragmatic arguments. I realize that Isaac may be tired of this ongoing controversy. But I try to console myself with the thought that, as they say, *amicus Plato* . . .

As I understand it, a pragmatic argument for a principle, *P*, is an argument that appeals to the desirable/undesirable consequences of *P*'s satisfaction/ violation. Here, my focus is on pragmatic arguments for various "rationality constraints" on a decision maker's state of mind: on his beliefs or preferences. An argument of this kind purports to show that a violator of a given constraint can be exposed to a decision problem in which he will act to his guaranteed disadvantage. To put it dramatically, as such arguments frequently are put, a violator of a constraint can be exploited by a clever bookie, who in order to set up his exploitation scheme doesn't need to know more than the agent who is being exploited. The locus classicus for such arguments is the pronouncement by Frank Ramsey: "If anyone's mental condition violated these laws [= the laws of probability], . . . [h]e could have a book made against him by a cunning bettor and would then stand to lose in any event" (Ramsey 1926, p. 78). Examples of pragmatic arguments of this kind are synchronic Dutch books for the standard probability axioms, diachronic Dutch books for the more controversial principles of reflection and conditionalization, and money pumps for the transitivity requirement on preferences.

When one examines various examples of such pragmatic arguments, one thing is especially striking. The proposed exploitation setups share a common feature. Suppose that the violator of a given constraint is logically and mathematically competent. Assume also that he prefers being better off to being worse off and that he acts accordingly. Then, it turns out, he can be exploited *only if* he is *disunified* in his decision making. By this I mean, roughly, that exploitation is possible only if the agent makes decisions on various issues he confronts one by one, rather than on all of them together. Instead of deciding on the whole package, the agent proceeds in a piecemeal fashion and decides on each component of the package separately.[1]

An agent can be disunified in this sense either synchronically or diachronically. In the synchronic case, he is presented with a number of opportunities, each of which he can accept or reject. He proceeds to make a number of choices, one for each of the opportunities in question. Unified decision making would instead involve considering all these opportunities together, followed by a joint choice of a particular *configuration* of opportunities that the agent is willing to accept. In the diachronic case, the opportunities are expected to arise at different points in time, and a disunified agent defers his choice with respect to each opportunity to the time at which it will be offered. A unified approach would instead involve one decision on the whole package of opportunities, that is, a joint choice of a particular configuration of opportunities, present and future. Thereby, the need for piecemeal choices is preempted.

[1] To avoid possible misunderstandings, let me stress that the disunification I have in mind is not a form of schizophrenia. It is not that one "part" of the agent decides on one issue and another "part" on another issue. The disunification is not in the subject but in the object of decision making: Different issues are addressed by the agent separately rather than together.

To avoid a very different kind of misunderstanding, I should add that the convention I have adopted in this chapter – to treat the violator of a constraint as a male and the bookie as a female – is motivated by considerations of simplicity.

This sort of unity in decision making may be quite costly and is often inconvenient, especially when it concerns opportunity packages that are spread over time. For various reasons, we tend to find it easier to deal with different issues separately, rather than in a wholesale manner. In diachronic cases, there is an additional difficulty of limitations in control; we may be unable to influence our future behavior. Now, as I have suggested, the exploitation setups described in the pragmatic arguments for various constraints on beliefs and preferences work only for the agents who not only violate these constraints but also are disunified in decision making. Consequently, these arguments should be seen as at most delivering *conditional* conclusions: "If you want to afford being disunified as a decision maker, then you'd better satisfy these constraints." (Even these conditional conclusions are at most pro tanto. Vulnerability to exploitation should be avoided, but not at all costs.) Arguments of this kind fail to establish the inherent rationality of the constraints under consideration. To make categorical claims of rationality, one would need to argue for these constraints in some other way. In fact, I believe that some of the constraints for which pragmatic arguments have been provided, such as the principle of reflection, are by no means inherently rationally required. Other constraints, on the other hand, such as transitivity of preference or standard probability axioms, do seem to have a strong claim for being rationality canons. (On this issue, I think, Levi and I are much in agreement.) But, on the view I would like to defend, there still is something to be said for pragmatic arguments, both of synchronic and of diachronic variety, if they are interpreted in the way suggested above: as arguments for various conditions that level the ground for disunification in decision making.

In this chapter, I do not try to provide a conclusive defense of this interpretation of pragmatic arguments. What I do, however, is to provide several illustrations of the connection between exploitability and disunification.

Note also that diachronic arguments on my reading come out as being *stronger* than the synchronic ones, for the following reason: Decision unification, as a rule, is less easy to achieve diachronically than synchronically.

Levi's view of the status of pragmatic arguments is not at all like mine (cf. Levi 2002). In a way, it is directly opposed to my position. According to him, only synchronic pragmatic arguments are valid. The diachronic ones, he argues, lack validity. However, before I explain why he takes this view, let me present some examples of arguments of both kinds, in order to provide a background for the discussion. In these examples, I rehearse some material that will be familiar to many readers. I deplore this, but I fear that the ensuing discussion otherwise would be too abstract. Such a reminder might help to put some meat on the bare bones of disagreement.

1. A SYNCHRONIC DUTCH BOOK ARGUMENT FOR PROBABILITY LAWS

In this argument, it is assumed that an agent's probability assignments – his degrees of belief – are his guides to action. As such, these assignments are related to the agent's betting dispositions or, better put, to his commitments to betting behavior. The agent who assigns a probability for a proposition is committed to a specific *betting rate* for the proposition in question.

Consider a bet on a proposition A that costs C to buy and pays S if won. (S and C are monetary amounts.) S is the *stake* of the bet (the prize to be won), while C is its *price*. A bet shall be said to be *fair* if the agent is prepared to take each of its sides: to buy it or to sell it, whatever he is being asked to do. To pronounce a bet as fair, for a given agent, is thus to ascribe to the agent a commitment to a certain betting behavior. Suppose that for various fair bets on A, with different stakes and prices, the *ratio* between the stake and the price remains constant. If the stake is increased or decreased, the price of a fair bet is increased or decreased in the same proportion. This simplifying assumption (which would follow if we supposed that the agent is seeking to maximize his expected monetary payoff) is reasonable at least within a certain range, when the monetary amounts S and C are not too high. Within that range, the assumption of the constant ratio for all fair bets on a given proposition is not very problematic, for in those cases we may assume that utility is proportional to money.

We shall call this constant ratio the *betting rate* for A. That is, the *betting rate* for A is the quotient C/S for a fair bet on A. The agent's *probability* for A, $P(A)$, is identified with his betting rate.[2] Probabilities (= degrees of belief) can be measured by betting rates if the agent's betting commitments are determined by his probability assignments.[3]

[2] Given this interpretation of probabilities as betting rates, the expected value of buying a fair bet on A with prize C and stake S is zero:

$$[P(A) \times S] - C = [C/S \times S] - C = 0.$$

Similarly, selling such a fair bet has an expected value equal to zero:

$$C - [P(A) \times S] = C - [C/S \times S] = 0.$$

[3] There is a troublesome assumption lurking behind this approach to probabilities in terms of fair bets. It is by no means obvious that fair bets exist in the first place. The agent's highest buying price for a bet on A with a given stake might well differ from his lowest selling price for that same bet. The highest price he is willing to pay for a bet on A may well be lower than the lowest price for which he is willing to sell it. If this is the case for all stakes, then for no bet on A the agent would be willing to take both of its sides. If the constancy assumption still holds (for small stakes at least), the agent will have two betting rates

Example: If a bet on A with a stake $S = \$20$ and a price $C = \$9$ is fair for the agent, then his betting rate for A equals $9/20 = .45$, which means that we can set the agent's probability for A as equal to .45.

A *Dutch book* is a system of bet offers on various propositions that, if taken, would give the agent a positive loss whatever happens. A *synchronic* Dutch book is a system of simultaneous bet offers of this kind. In a *diachronic* Dutch book, the offers are made at different points in time.

If the agent violates standard probability laws, he is vulnerable to a synchronic Dutch book. It can be shown that there exists a set of bets on various propositions such that each of the bets is fair but together they would give the agent a guaranteed loss. This provides a pragmatic argument for obeying the probability laws.

Example: According to the addition axiom for probabilities,

If propositions A and B are logically incompatible, then $P(A \vee B) = P(A) + P(B)$.

Suppose the agent's probability assignments violate this axiom. For example, suppose that $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, while $P(A \vee B) = \frac{3}{4}$.

We sell to the agent bets on A and on B, respectively, each with a stake S and a price $\frac{1}{2}S$; and we *buy* from him a bet on the disjunction A or B with the same stake and a price $\frac{3}{4}S$. Given his probabilities, all these bets are fair. Our guaranteed profit is $\frac{1}{4}S$ (see Table 19.1).[4]

It is easy to see that the violator of the addition axiom is being exploited in this setup only because his decision making is disunified: He decides on each bet separately, rather than on all the three bets together. If the agent did

for A, one for the bets he buys (specifying the highest price-stake ratio he is willing to accept as a buyer) and the other for the bets he sells (specifying the lowest such ratio he is willing to accept as a seller). This would mean that probabilities could not be measured by betting rates, but it might still be possible to think of them as *partial* determinants of the agent's betting dispositions: The agent's probability for A would lie between his selling betting rate and his buying betting rate. (Another alternative would be to let the agent's probability be a pair of numbers, rather than a single number.) This would create problems for Dutch book arguments (see next note).

[4] If the agent's buying betting rate and his selling betting rate need not coincide, then this example only shows how one can exploit an agent whose selling rate for $A \vee B$ is lower than the sum of his buying rates for A and for B. If – as is reasonable to assume – buying rates do not exceed selling rates and if probabilities are assumed to lie between the former and the latter, then an agent such as this violates the addition axiom for probabilities. But the opposite does not hold: The agent's probability for $A \vee$ B may be lower than the sum of his probabilities for A and for B, but his selling rate for the disjunction need not be lower than the sum of his buying rates for the disjuncts. Such an agent will not be exploited by the setup we have described. The same remarks apply, mutatis mutandis, to the diachronic exploitation scheme that is described in the next section.

Table 19.1: *The agent's gains and losses*

Possibilities	Bet on A (bought)	Bet on B (bought)	Bet on $A \vee B$ (sold)	Total
A	$S - \frac{1}{2}S$	$-\frac{1}{2}S$	$\frac{3}{4}S - S$	$-\frac{1}{4}S$
B	$-\frac{1}{2}S$	$S - \frac{1}{2}S$	$\frac{3}{4}S - S$	$-\frac{1}{4}S$
$\neg(A \vee B)$	$-\frac{1}{2}S$	$-\frac{1}{2}S$	$\frac{3}{4}S$	$-\frac{1}{4}S$

the latter, then – assuming he is logically and mathematically competent – he would certainly choose not to accept the whole bet package, since a simple calculation would show that refusing all three bets would be better for him, whatever happens. Of course, in this unified mode, the agent might still decide to accept two bet offers out of three, say, to buy the bet on A and the bet on B, but this would not lead him to a sure loss.[5]

2. A DIACHRONIC DUTCH BOOK ARGUMENT FOR THE REFLECTION PRINCIPLE

This principle expresses a requirement that current probability assignments should reflect one's expectations concerning one's future probabilities.[6] Thus, in particular, my current probability for a proposition A conditional on the supposition that my future probability for A will at most be k should itself at most be equal to k.

Principle of Reflection: $P(A/P'(A) \leq k) \leq k$, provided that $P(P'(A) \leq k) > 0$,

where P is the agent's current probability, at time t, and P' is his probability assignment at an arbitrary future point of time $t'(t' \geq t)$.

It is a standard objection to Reflection that this principle requires the agent to have an unlimited trust in his own future cognitive abilities. If the agent lacks

[5] In Skyrms's "Higher Order Degrees of Belief" (1980), it is suggested that an agent who is vulnerable to such a predicament must be logically confused. Such an agent seems to evaluate one and the same betting arrangement differently depending how it is presented to him or her: as a set of three bets or as one composite bet that guarantees a net loss, whatever happens. However, this suggestion can't be right, as far as I can see. It is true that the agent we consider views each bet in the set as agreeable and yet assigns negative value to the bet package as a whole. This does not mean, however, that he or she evaluates the betting setup differently under different logically equivalent descriptions. It means only that his or her valuations are not additive: He or she ascribes to the whole package a value that is not the sum of the values of its parts. But one need not be logically confused in order to have nonadditive valuations (cf. Schick 1986).

[6] This section and the next one draw on Rabinowicz 2000.

this trust, he might well violate Reflection. Having good grounds to doubt one's cognitive rationality in the future might make it cognitively rational *now* to have nonreflective probability assignments. To take a simple case, suppose the agent has some grounds to believe that his future probability for A, at $t' > t$, might be too low as compared with the evidence he will have available at that time. To take an extreme case, suppose the agent expects to be subjected to a brainwash that will at t' make him unreasonably skeptical about A. Then, his present conditional probability for A on the hypothesis that $P'(A) \leq k$, where k is low, should be higher than k. Clearly, a brainwash is just an example. Any kind of predicted cognitive deterioration will do.

Still, as has been shown by van Fraassen (1984), an agent whose probability assignments violate reflection is vulnerable to a diachronic Dutch book, quite independently of whether these violations of Reflection are well grounded or not. That is, we can set up a system of bets, to be offered to the agent at various times (t and t'), such that (i) each bet, when offered, is advantageous from his point of view at that time, but (ii) together, they guarantee him a certain loss.

Here is an example (cf. Christensen 1991). Suppose that an agent's probability assignment P at t violates Reflection:

(i) $P\left(A/P'(A) \leq \frac{1}{2}\right) = \frac{3}{4}$.

Let E stand for the proposition that $P'(A) \leq \frac{1}{2}$, and suppose that

(ii) $P(E) = \frac{1}{5}$.

At t, a bookmaker offers the agent two bets:

(1) a bet on E that costs one unit and pays five if won;
(2) a *conditional* bet on A given E, that costs fifteen and pays twenty if won.

In a conditional bet, the price of the bet is refunded if the condition is not realized. It is easy to see that both bets are fair, in terms of the agent's probabilities at t. (For conditional bets, the fair betting rate equals the agent's conditional probability.)

Then, at t', if E will be the case, that is, if the agent's probability for A will not exceed $\frac{1}{2}$ at t', the bookmaker will offer to *buy* from the agent a third bet:

(3) a bet on A that costs ten and pays twenty if won.

Table 19.2: *The agent's gains and losses*

Possibilities	Bet on E bought at t	Bet on A given E bought at t	Bet on A sold at t'	Total
$E \wedge A$	$5 - 1$	$20 - 15$	$10 - 20$	-1
$E \wedge \neg A$	$5 - 1$	-15	10	-1
$\neg E$	-1	Called off	–	-1

Selling this bet will be fair or "more than fair" in terms of the agent's probabilities at t' if E will then be the case. Under this assumption, the agent's probability for A at t' will at most be equal to $\frac{1}{2}$.

If the agent accepts all these bet offers, then he will lose one unit, whatever happens. The reason is simple: If E won't be the case, the agent will lose the bet on E. The conditional bet on A given E will be then be off and no bet offer will be made at t'. On the other hand, if E will be the case, the agent will win the bet on E but then the bookie will be able to buy back from him the bet on A *at a lower price*. Since this price difference $(15 - 10 = 5)$ exceeds the net gain from the bet on E $(5 - 1 = 4)$, the agent will again suffer a total loss (see Table 19.2).

There is an obvious objection to this line of reasoning. A pragmatic argument for a formal constraint on the agent's probability assignments is supposed to demonstrate that violating this constraint would be to the agent's disadvantage by his own lights – in the light of the information that stands at his own disposal. To be effective, such an argument should therefore be based on the assumption that the agent who is to be exploited knows at least as much as his would-be exploiter. Insofar as the exploiter acts on a definite plan of action, this plan must therefore be known to the agent. Thus, the agent must have *foresight*.

But, surely, the objection continues, if the agent has foresight and thus knows what is kept in store for him at t' if E will be the case, he can at t stop the Dutch book from the start by simply refusing to take the earlier bets. The agent will thereby frustrate the bookmaker's plans to exploit him and the whole book will crumble. By refusing to accept the earlier bets, the agent will avoid being asked to sell the bet on A in the future. The agent will avoid being exposed to an opportunity that he would then be willing to accept but that he now – by his *present* lights – finds unattractive (cf. Levi 1988 and Maher 1992).

Here is how this objection could be met (cf. Skyrms 1993): Suppose the bookmaker is *persistent* in her scheme and the agent knows this. What we mean by persistency is that later bet offers are not conditioned on the acceptance of earlier bet offers. More precisely, at t', if E will then be the case, the bookmaker will offer to buy the bet on A *come what may*. She will make this offer even if the agent were to refuse to buy any bets at t. Suppose also that the bookmaker makes all the three bets "more than fair": For each bet offer he accepts, the agent gets a small reward ε. (Assume, however, that $3\varepsilon < 1$. Then the extra rewards still won't compensate the agent for his total loss if he accepts all the three offers.)

In terms of his probabilities at t, selling the bet on A is unattractive for the agent. Were that opportunity offered at t, conditionally on E, he would never accept it. (At t, the expected value of that transaction equals $10 - (\frac{3}{4} \times 20) + \varepsilon = -5 + \varepsilon$.) At t', however, if E will be the case, selling the bet on A will become attractive given the agent's new probabilities. Realizing this beforehand, the agent can use backward-induction reasoning to solve the problem. Insofar as he has trust in his future practical rationality, the agent can at t predict he will sell the bet on A at t', if E will then be the case, in view of the attractiveness of that trade at that time.[7] The agent will sell it, whatever bets he might have accepted earlier on. But then buying the conditional bet on A given E at t doesn't make things worse in any way. In fact, it makes them better, by an extra ε. Similarly for the bet on E, and even more so for these two bets taken together (which improves the agent's prospects by 2ε). Thus, if the bookmaker is known to be persistent, the backward-induction reasoning leads the agent to accept all the bets that are offered, at t and t', even though he knows this will result in a certain loss.[8]

[7] Backward-induction reasoning is based on a trust in one's future rationality. The agent assumes that he is going to act rationally in the future, which allows him to predict his future behavior and then make his current choices in the light of these predictions. One might therefore wonder whether this kind of reasoning is available to agents who violate Reflection because they expect to become cognitively irrational in the future. The answer is that a violator of Reflection can make use of backward induction as long as he expects to remain *practically* rational, that is, rational in what he does given what he believes and values. That his future *beliefs* need not be rational as judged by his present lights is another matter.

[8] The assumption of persistency on the part of the bookmaker is never explicitly stated in van Fraassen's "Belief and the Will" (1984). Nor is it emphasized in the well-known diachronic Dutch book argument for conditionalization, due to David Lewis (cf. Teller 1973). As a result, Levi in his "Demons of Decision" (1988) is able to argue that a diachronic Dutch book can easily be avoided by refusing to accept the bets that are offered at the initial stage. As it is clear from Levi's discussion, he thinks such a refusal would let the agent off the hook by preventing future betting offers that are so disadvantageous by his present lights (cf. Levi's discussion of Case 2 on pp. 204f). Levi's idea is made more explicit in Maher 1992.

As a matter of fact, using backward induction is something of an overkill in this case. A simple dominance reasoning is enough. For the agent to conclude that there is no reason to abstain from the bets offered at t, he need not assume he will do the rational thing at t'. It is enough if two conditions are met. (i) The agent believes his actions at t won't influence the bet offer at t', if E will be the case. (ii) The agent expects to deal with that offer at t' in the same way, whatever he might do at t. As (i) and (ii) imply that the agent's present actions won't influence his actions in the future nor their effects, he can conclude that buying bets at t is preferable to abstaining, as it improves the agent's prospects by 2ε independently of what he will decide to do at t'.

As in the example in the previous section, it is easy to see that the violator of reflection is being exploited in this setup only because his decision making is disunified: He decides on each bet separately, when it is being offered, rather than on all the three bets together. If the latter was done, then – assuming the agent is logically and mathematically competent – he would certainly choose not to accept the whole bet package, since a simple calculation would show that refusing the three bets would be better for him, whatever happens. The salient feature of this case is the agent's disunification over time: Even if he were synchronically unified and thus made a joint decision on the two bets offered to him at t, the agent would still be exploited as long as his decision on the bet offered at t' is left to that future occasion. The two bets offered at t, if considered together, promise the agent a positive expected profit (of 2ε),[9] and thus represent together an attractive opportunity. And the same applies to the third bet, if and when it is offered at t': Its expected value at that time is positive. Together, however, these two opportunities guarantee a sure loss.

That the persistency of the exploiter closes this gap in van Fraassen's and Lewis's arguments has been shown in Skyrms 1993, in his comments on Maher's paper. As Skyrms puts it: "Why is it assumed [by Maher and Levi] that the cunning bettor will just go home if [the agent] refuses to bet today?... Deciding not to bet ever is not an option.... Even though [the agent] will see it coming, she will prefer the sure loss ... because doing so looks strictly better to her than the alternative" (ibid., pp. 323f.). And he concludes: "Seeing it coming does not help" (p. 326).

[9] Here, we assume that the agent satisfies the standard probability axioms. Given this assumption, his expected value if he accepts both bets offered at t equals

$$\begin{aligned}
&P(E \wedge A) \times (5 - 1 + 20 - 15 + 2\varepsilon) + P(E \wedge \neg A) \times (5 - 1 - 15 + 2\varepsilon)\\
&\quad + P(\neg E) \times (-1 + 2\varepsilon)\\
&= 3/20 \times (9 + 2\varepsilon) + 1/20 \times (-11 + 2\varepsilon) + 4/5 \times (-1 + 2\varepsilon) = 2\varepsilon.
\end{aligned}$$

3. MONEY PUMPS AGAINST AGENTS WITH CYCLICAL PREFERENCES

Suppose the agent's preferences with respect to alternative outcomes x, y, and z are cyclical: $x \prec y \prec z \prec x$ ($\prec$ here represents his preference relation).[10] Let x be the status quo alternative that will be realized if no action is taken by

[10] The simplest case of cyclicity arises when the different aspects or respects of comparison are aggregated by means of a *Pareto rule*: *If* y is strictly preferred to x in some respect and x and y are equipreferred in all other relevant respects, *then* $x \prec y$. To state that y is weakly preferred to x in respect a, where a is one of the relevant aspects of comparison, let us use the notation $x \preceq^a y$. We define aspectual equipreference and strict preference in the standard way:

$$x \approx^a y = (x \preceq^a y) \text{ and } (y \preceq^a x).$$
$$x \prec^a y = (x \preceq^a y) \text{ and not } (y \preceq^a x).$$

Suppose we allow the aspectual equipreference to be nontransitive. The following situation then becomes possible:

$$x \approx^a y, \quad y \approx^a z, \quad \text{but } z \prec^a x.$$

Equipreference with respect to a is nontransitive if the differences with respect to that aspect between pairs of alternatives, say, between x and y and between y and z, are either imperceptible or simply unimportant in the sense of being too small to matter. But when we compound such differences, by first moving from x to y and then from y to z, the compound difference becomes perceivable, or sufficiently large to matter.

While aspectual preferences do not involve cycles in strict preference, such cyclicity can arise in the preference-all-things-considered *if* we aggregate aspectual orderings by means of the Pareto rule. Then, if we let $\{a, b, c\}$ be the set of all the relevant aspects of comparison, the following example shows that $\prec$ can be cyclical:

$$\begin{array}{lll} x \approx^a y, & y \approx^a z, & \text{but } z \prec^a x. \\ y \approx^b z, & z \approx^b x, & \text{but } x \prec^b y. \\ z \approx^c x, & x \approx^c y, & \text{but } y \prec^c z. \end{array}$$

Aspect b resolves the tie between x and y to y's advantage: $x \prec y$; similarly, the tie between y and z is resolved to z's advantage by aspect $c : y \prec z$; and, finally, the tie between x and z is resolved to x's advantage by aspect a: $z \prec x$.

Schumm's well-known example has this preference structure (cf. Schumm 1986). There are three boxes, x, y, and z, with Christmas tree ornaments. Each box contains three balls: one red, one blue, and one green. But for each color, the balls in different boxes exhibit that color in differing shades, more or less pleasing to the eye. $\preceq^a$ reflects the agent's preferences over the boxes with respect to the shade of their red balls, while $\preceq^b$ and $\preceq^c$ reflect his or her preferences with respect to the shade of the blue and the green balls, respectively. Closely similar shades of the same color are indiscernible to the eye, but the agent can still discern between the shades that are sufficiently different. If the agent prefers some shades to others when he can discern between them, but not otherwise, his preference structure $\{\preceq^a, \preceq^b, \preceq^c\}$ may be like the one specified above. Then we get cyclicity in the agent's all-things-considered preference if we use the Pareto rule for aggregation.

the agent. He is offered an alternative y in exchange for x. The exchange costs a small amount ε that does not reverse the agent's preference for y over x. After the agent has made the exchange, he is offered to trade y for a preferred alternative z, if he pays an additional ε. After he has made this second exchange, the agent is offered to trade z for a preferred alternative x, if he again pays ε. After the three exchanges, the agent is back to where he started, minus 3ε. He has been used as a money pump. Isn't it irrational to be vulnerable to such a predicament?[11]

In *indefinite* money pumps, the process of exchange continues until the agent is ruined. Here, we consider only *finite* pumps, where the exploitation stops after k full rounds. For simplicity, assume that $k = 1$. That is, the pump stops after three exchanges. For this short pump to work, the extra payment of ε should not reverse the agent's preferences at any stage, at least up to 3ε. Thus, we need to assume that

$$x \prec y - \varepsilon \prec z - 2\varepsilon \prec x - 3\varepsilon.$$

The money pump argument, as described above, confronts an obvious objection: The pump seems to work only if the agent doesn't know he is being taken for a ride. He would refuse to trade if he knew what's kept in store if the exchange is made.[12] The point of the objection is that the important precondition of foresight, which should be satisfied by diachronic pragmatic arguments, is not satisfied in the money pump in its traditional version.

The common view is therefore that a prudent agent with foresight would simply refuse to be pumped, because he would see what's coming. The agent would realize that the first trade would lead to the second trade, which would lead to the third trade, which would get him back to where he started, minus the payments. At some point, therefore, before completing the full circle, he would refuse to make any further trades.

This idea of foresight prudently employed as a defense against exploitation can be made more precise in different ways. Here is one. When an agent is

[11] Cf. Davidson, McKinsey, and Suppes 1955 or Raiffa 1968.

In fact, in order to set up a money pump against the agent, his preferences over basic alternatives need not be strictly cyclical. It is enough if the agent's equipreference ordering is nontransitive: $x \approx y$, $y \approx z$, but $z \prec x$. If the agent is prepared to pay ε for an exchange from z to x, then we can use a small reward δ to get him to exchange the status quo alternative x for y, which the agent prefers just as much. Then we offer the agent another δ if he trades y for z, and we finally let the agent pay ε if he trades z for x. If δ is small enough, so that $2\delta < \varepsilon$, the agent will end up with less money than he has started with.

[12] Cf. Schick 1986 and Schwartz 1986.

fully informed about his sequential choice problem, then – if he is rational and has a robust trust in his future rationality – he can solve the problem using backward induction. The agent can first determine which move would be rational to make at the last choice node of each branch of the decision tree, when it is clear which payoffs would result from each move. Relying on his own future rationality, the agent predicts that he would make that rational move if he were to reach the node in question. Taking his trust in his future rationality to be robust, the agent expects to hold on to these predictions on reaching the penultimate choice node on each branch. This allows the agent to determine ex ante which move would be rational at each such penultimate node and thus, again relying on his future rationality, to predict his own behavior at that node. Continuing in this way, from the end points of the tree to its beginning, such a sophisticated chooser finds out which moves are rational at each choice node of the tree. At each node, the move prescribed by backward induction is the one that would be optimal on the assumption that any move made at that node would be followed by backward induction moves at all the later nodes.

As has been argued by Ned McClennen in *Rationality and Dynamic Choice* (1990, sec. 10.2), a sophisticated chooser is not vulnerable to a money pump. I have argued for the same claim myself in "To Have One's Cake and Eat It, Too" (1995). As McClennen's argument contains some minor mistakes, the presentation below follows my 1995 paper.

As before, we assume that the agent's preferences with respect to x, y, and z are cyclical and that his cyclic preferences are not reversed by extra payments. We also assume that $x - 3\varepsilon \prec x$, which means that the agent who starts with x and ends up with $x - x - 3\varepsilon$ will suffer a definite loss from his own point of view. In fact, we take it that $x - 3\varepsilon$ is dispreferred by the agent not just to x, but also to any alternative he prefers to x. Thus, in particular, since $x \prec y - \varepsilon$, it also holds that $x - 3\varepsilon \prec y - \varepsilon$.

We now consider the agent's sequential choice problem that consists of three trade offers, as shown in Figure 19.1. The forks in this tree are the agent's choice nodes. Going up means accepting an exchange, going down is rejecting it. At the end point of each branch in the tree, the final outcome is specified. The status quo alternative is x, which means that x will obtain if the agent at the starting point refuses to trade, that is, goes down in the first node. If the agent instead goes up, that is, makes one exchange, but then stops, he ends up with $y - \varepsilon$. If the agent makes two exchanges and then stops, he ends up with $z - 2\varepsilon$. Finally, if the agent makes all the three trades, he arrives at $x - 3\varepsilon$.

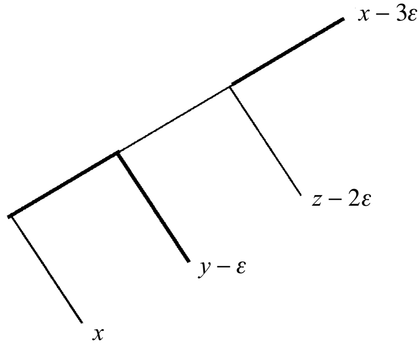


Figure 19.1: Money pump.

The bold lines in the tree represent backward-induction moves. At the third node, the agent's preferences dictate trading, since he prefers $x - 3\varepsilon$ to $z - 2\varepsilon$. Given that the agent expects to trade at the third node were he to trade at the second node,[13] his choice at the second choice node should be to *refuse* to trade: This refusal gives him $y - \varepsilon$, which is clearly preferred to $x - 3\varepsilon$. But then, given that he expects to refuse at the second node, the choice at the first node should be to *trade*, since he or she prefers $y - \varepsilon 0$. Thus, the sophisticated chooser will make just one exchange and then stop. Even though he preferences are cyclical, he will not be pumped.

A pump like the one described above may in general involve any number n of cycling basic alternatives, $x_1, \ldots, x_n$ ($n = 3$ in our example, in which the basic alternatives are *x, y*, and *z*) and any number *k* of full rounds (in our example, $k = 1$). It is easy to see that, for any such pump, the sophisticated chooser will never end up with an alternative he disprefers to the status quo alternative (= the alternative he has started with). The reason is simple: If he trades in the first move, then he correctly expects this move to be followed by a series of moves dictated by backward induction. Thus, the agent has a definite and correct expectation as to the final outcome of his trading move. If he nevertheless does trade, he must prefer this outcome to the status quo alternative. It follows then that he either refuses to trade at all or, if he trades, the final result is an outcome he prefers to the one he has started with.[14]

[13] We assume that the agent (i) knows his own preferences and (ii) expects them to remain unchanged as he moves along the decision tree. Thus, if the agent now prefers $x - 3\varepsilon$ to $z - 2\varepsilon$, then he expects to have this preference in each possible future decision node.

[14] More precisely, it can be proved that the agent will stop the pump at some point before the completion of the first round. For the proof, see Rabinowicz 2000.

There is a well-known objection to backward induction: This procedure presupposes that the agent, at each nonterminal choice node *c* that he could reach, would assume that any move at that node would be followed by backward-induction moves at all the subsequent choice nodes. But suppose that *c* is one of those nonterminal choice nodes in the tree that can be reached only by a series of moves that themselves are forbidden by backward induction. Wouldn't the agent at *c* then have grounds to doubt whether he would act in accordance with backward induction at the subsequent choice nodes? If he didn't do it earlier, why expect he would do it later? To put it differently: If, for the argument's sake, backward induction is supposed to codify rational behavior, how can it rely on the presupposition that the agent's trust in his own future rationality is robust? How can the agent be supposed to continue to have this trust, whatever evidence he might accumulate about past behavior? But without such presupposition of continued trust, backward-induction reasoning does not go through.[15]

This objection does not apply to the short money pump described above. There, it is only the terminal choice node that cannot be reached without violation of backward induction. But what is rational at the terminal choice node does not depend on what the agent expects to do in the future. On the other hand, at the nonterminal choice nodes in this decision problem, the agent lacks evidence about prior violations.

The objection does apply to more complicated money pumps, which involve several rounds or are based on cycles consisting of more than three alternatives. Still, if a money pump is not too long, and if the sophisticated agent starts out with a firm conviction about his commitment to the backward-induction procedure, the evidence about his deviant past behavior might never be extensive enough to shatter his initial conviction. The agent will be able to explain away his past deviations from the backward-induction path as isolated mistakes that would not be repeated in the future.[16]

Are we then out of the woods, at least as far as relatively short sequential decision problems are concerned? Is foresight, coupled with sophistication, sufficient to save an agent with cyclical preferences from being pumped? Not quite, I am afraid. What follows is a description of a money pump that can be used against a sophisticated chooser (cf. Rabinowicz 2000).

[15] Cf. Binmore 1987; Reny 1988, 1989; Bicchieri, 1989; and Pettit and Sugden 1989. For various defenses of backward induction against this objection, either in general or for a limited class of cases, see Sobel 1993; Aumann 1995, 1998; Rabinowicz 1998; and Broome and Rabinowicz 1999.

[16] These remarks also apply to the modified money pump that I am going to consider below.

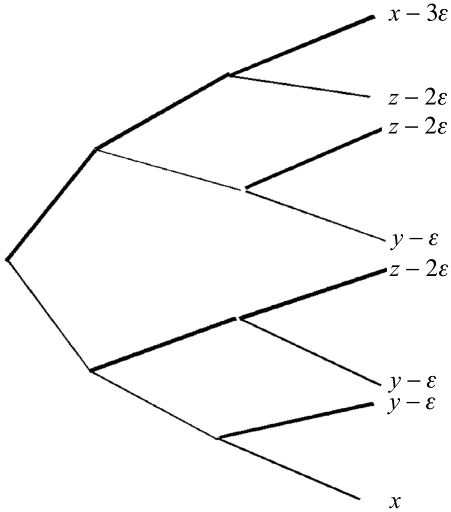


Figure 19.2: Money pump with persistent offers.

In the money pumps discussed up to now, the series of trades terminates as soon as the agent refuses to make yet another exchange. No further trade offers are forthcoming. Suppose we change this feature of the decision problem. The would-be exploiter is now assumed to be *persistent*: If you refuse a trade offer, she comes back with the same offer at the next stage.[17] There are three stages at which offers are made. The decision tree for the new money pump has the form shown in Figure 19.2. As before, trades and refusals to trade are represented as upward and downward moves, respectively. If the agent each time refuses to trade, he ends up with x. If he trades just once (which may be done at any stage), he ends up with $y - \varepsilon$. If he trades twice (which, again, may be done at any two stages), he ends up with $z - 2\varepsilon$. Finally, if the agent makes all the three trades, he ends up with $x - 3\varepsilon$, that is, returns to where he has started minus extra payments.

The bold lines stand for the moves prescribed by backward induction. Thus, at each terminal choice node, backward induction prescribes trading: The trades give the agent his preferred alternative, and he knows that no new trade offers will be forthcoming. Since the agent predicts he will trade at each

[17] Obviously, it is a variant of the same idea that was exploited by Skyrms in his treatment of diachronic Dutch books (cf. the previous section).

terminal choice node, he should trade at each penultimate node as well. In the upper penultimate node, the agent predicts that trading would eventually lead to $x - 3\varepsilon$, while refusal would lead to $z - 2\varepsilon$, which he disprefers. Analogously, at the lower penultimate choice node, the agent predicts that trading would eventually lead to $z - 2\varepsilon$, while refusal would lead to $y - \varepsilon$, which he disprefers. Given that the agent predicts he will trade at each node after the first one, he should trade at the first node as well. Trading at that node would eventually lead to $x - 3\varepsilon$, while refusal is predicted to lead to $z - 2\varepsilon$, which he disprefers.

We conclude, then, that in this modified money pump, a sophisticated chooser with cyclical preferences will be pumped: He will trade each time, which will get him back to the status quo minus extra payments. The reason is obvious. The exploiter, being persistent in her offers, never lets the agent off the hook. Refusing to trade at an early stage does not terminate the pump: The trade offer will be repeated.[18]

That backward induction implies that continuous trading, if the would-be exploiter is persistent, is a robust result, which can be generalized to pumps with an arbitrary number of stages (for the proof, see Rabinowicz 2000). Such pumps may be based on any number n of basic cycling alternatives, $x_1, \ldots, x_n$ (in our example, $n = 3$), and they may involve any number k of full rounds (in our example, $k = 1$). What we need to assume to obtain this result is only that the small payment required by each trade would never reverse the agent's preference with regard to the basic cycling alternatives, independently of how many payments have already been made.

As in the examples in the two preceding sections, the agent with cyclical preferences is being exploited in this setup because his decision making is disunified. More precisely, the agent is disunified over time: He decides on each exchange separately, at the stage when it is being offered, rather than on all the three exchange stages together. If the agent did the latter, then, we may safely assume, he would certainly not choose to accept all the three

[18] Despite obvious similarities, there is an important difference between this modified money pump and Skyrms's exploitation setup for a violator of reflection. In Skyrms's setup, backward induction need not be used by the agent: Dominance is enough, as we have seen. In my money pump, however, dominance reasoning is inapplicable, for two reasons: (i) The agent's choices at earlier stages influence the opportunities he will confront later: Depending on whether the agent trades in a given stage, he will be offered either another trade or the same trade at the subsequent stage. (ii) The agent's decisions on whether to accept a given exchange crucially depend on his expectations about how he will deal with the future exchange opportunities.

exchanges, since a simple calculation would show that refusing the three bets would get him the same outcome x without any extra payments.[19]

4. LEVI'S CRITICISM OF SYNCHRONIC PRAGMATIC ARGUMENTS

In "Money Pumps and Diachronic Dutch Books" (2002), Levi considers the scenario of my money pump with persistent offers and of Skyrms's version of the Dutch book against violators of reflection. He argues that there exists a decisive difference between these exploitation setups and the ones that are being used in synchronic Dutch books. The difference has to do with the range of options that are available to the agent. In a synchronic setup, an agent who violates a certain constraint is shown to act in a way that is *dominated by an option that stands at his or her disposal*. Something must be deeply wrong with a person who acts like this. The person can't be rational if he or she could have opted for an action that would yield better results under all possible circumstances. Thus, in a synchronic Dutch book against a violator of the addition axiom for probabilities, the agent accepts all the three bets, even though he or she has at his disposal the option of refusing any bet in the package. As compared with accepting each bet, the option of refusing them all *dominates*: It would yield better results whatever happens.

By contrast, in a diachronic setup, think of the agent at the initial choice node. He "has no control then over what he will choose later. He can only

[19] As his preferences are cyclical, it is not determined by our description of the case what particular outcome *would be* chosen in such a situation. But the assumption that an agent's pairwise preferences are cyclical is compatible with the possibility that, for any set X of alternatives, the subset of the alternatives in X that the agent finds choiceworthy is nonempty. On this issue, see Rabinowicz 2000. Levi's treatment of this issue in his "Money Pumps and Diachronic Dutch Books" (2002) is somewhat different from mine. What he points out, however, is that in the cases of "unresolved conflict" between various aspects of comparison, cyclicity in preferences revealed in binary choices will not imply any cyclicity in the agent's "categorical" preferences, all things considered. Rather, the latter will involve various incomparabilities between the alternatives. Both of us agree that, in a money pump, it is the agent's preferences in binary choices that account for exploitation. But it is preferences all-things-considered that come into play when the agent is confronted with the choice from the whole set $\{x, y - \varepsilon, z - 2\varepsilon, x - 3\varepsilon\}$. We may safely assume, that – in such a choice – the agent will definitely reject $x - 3\varepsilon$. Which of the remaining three alternatives he might opt for is less clear. On Levi's treatment, in the case of an unresolved conflict between x, y, and z, these three alternatives are all mutually incomparable with each other in terms of the agent's preferences all-things-considered while $x - 3\varepsilon$, all things considered, is dispreferred to x. Consequently, whatever the agent would choose from that set, he would definitely not choose $x - 3\varepsilon$, since he could always do better by opting for x instead.

predict what he will do" (Levi 2002, p. 239). As a consequence, when he is exposed to my money pump, and ends up making the three exchanges, "[he] is not choosing [at any point] an option dominated by another *available as an option* to him" (ibid., p. 241, Levi's emphasis). In particular, at the initial choice node, the alternative of refusing to exchange at any of the three stages is not an option that stands at the agent's disposal. Because of this absence of a dominating option, he cannot be charged with irrationality.

To be sure, Levi writes, a money pump like mine shows that an agent with cyclical preferences can be taxed for having preferences of this kind. The extra payments he incurs to get back where he started may be seen as such a tax. If the agent's preferences weren't cyclical, he would not have to pay. But vulnerability to taxation is not irrationality. Levi concludes:

> Money Pump arguments were designed initially to show that individuals who violate certain canons of rationality will end up choosing options that are dominated by other options available to them just like synchronic arguments do. Showing that violating these canons is one way, that in the face of other assumptions, makes one vulnerable to taxation, is no substitute. Those who use money pump arguments to defend acyclicity of preference have failed to show that decision-makers who violate acyclicity are driven to choose dominated options. (ibid., pp. 241f.)

Levi's diagnosis of Skyrms's version of the diachronic Dutch book against a violator of reflection is exactly similar: If the exploiter weren't persistent, the agent in that setup could get off the hook by refusing at t to buy the conditional bet on A given E. This is the bet the exploiter otherwise would buy back from him at a cheaper price at t', if E would be the case at that time. But if the exploiter is persistent, then, in the presence of small premiums for each accepted bet offer, "accepting [the conditional bet] at the initial step is preferable to rejecting it even though X [the decision maker] is doomed to receive a net loss relative to the initial status quo." However, this vulnerability to a diachronic exploitation doesn't show that the agent is irrational in any sense. Again, the reason is to be found in the range of options that are at the agent's disposal. The agent lacks control over future choices; he can only predict what he will do. The agent cannot at t decide not to accede to the bookie's offer at t'. Consequently, he cannot at t decide to remain in the initial status quo and refuse all the bet offers from the bookie. But this means that the agent cannot be accused of acting in a way that is being dominated by some of the available options.

> According to Skyrms's scenario, X is worse off, no matter how X chooses, than X was in the initial status quo. If X has the option of remaining in the status quo position, X

should do so [rather than act as he or she does]. But by hypothesis *X* does not have this option. *X* is not rationally compelled to choose an option dominated by *other available options*. . . . Buying [the bet on *A* conditional on *E*] at the initial stage is not dominated by refusing to buy it at that stage. Since these are the only two options, where is the beef?" (Ibid., p. 247, Levi's emphasis)

5. MY RESPONSE

Indeed, where is the beef? Levi is quite right that, in my money pump and in Skyrms's setup, both of us have assumed that the agent at the initial stage cannot control what he will do in the future. Consequently, his course of action is not dominated by any of his options. Instead, it is dominated by a certain *sequence* of options, each of which is available to the agent at a certain time. Such a dominating sequence of options in our setups consists in continuously refusing each offer made by the bookie. But the times at which different options in the sequence are available are not the same, and the sequence as a whole is not an option for the agent, at any time.

Now, I cannot speak for Skyrms, but as for myself, I assumed these limitations in the agent's diachronic self-control in order to make his exploitation come out as something unavoidable. I thought the diachronic case was in this way more worrisome than the synchronic one. In the latter, it certainly would be extremely unrealistic to suppose that the agent can separately decide on each of the bet opportunities that are on the table at a given time, but cannot decide jointly on the whole set of these opportunities.

However, in order to deal with the issue raised by Levi, we can simply modify the diachronic setup in such a way as to put the two kinds of arguments, the synchronic and the diachronic kind, on equal footing. Let us therefore assume that the agent, at the initial stage, does have an option to decide on the whole temporal sequence of his actions. But this agent never deliberates on the sequence as whole; he engages in a disunified decision making in which various bet offers are decided on separately, at the time when they are made. However, if the agent instead viewed the decision problem in a unified way, which he could, his decisions would make impact on his future behavior.

In this way, it seems, the synchronic and the diachronic exploitation setups become analogous. In the synchronic scheme the agent is also assumed to engage in a disunified decision making: He makes decisions on each bet separately. (Otherwise, no exploitation would take place.) But if the agent viewed the situation in a unified way, he would decide on the whole package of bets. The agent would then decide, jointly, which bets to accept and which to reject. It is in this sense that he has at his disposal an option of refusing all the bets, which dominates the way he actually behaves. This option is available, since it would figure in the agent's deliberation as one of the alternatives if he was unified and nothing hinders him from being so. After the modification, the diachronic setup is structurally similar in all these respects to the synchronic one.

In the synchronic setup, the option to refuse all the bets does not make it irrational for the agent to decide to accept any particular bet, when he asks himself whether to accept it or refuse it. When the agent asks this question, he engages in a disunified mode of decision making in which the option of joint refusal does not appear as one of the alternatives to be considered. The same applies to the diachronic case. In the diachronic setup, the mere presence of the option to refuse all the trade offers, current and future, does not make it irrational for the agent to accept any particular trade offer in the exploitation sequence, when he asks himself whether to accept that offer or to refuse it. For, again, when this question is asked, the agent is in a disunified mode of decision making in which the option of the wholesale refusal does not figure as one of the alternatives.

One might perhaps argue that there is this difference between the synchronic and the diachronic case: In the diachronic case, when I consider each trade separately, I use expectations of my future choices and my knowledge of past decisions in order to determine what will be the final outcome of the trade under consideration. In the synchronic case, however, when considering whether to accept a particular bet offer, I don't make any predictions about the decisions I take about other bets that are on offer. As long as each of these other bets still is under my deliberation, I can't relate to them in a predictive mode. At least on one interpretation of Levi's thesis that deliberation crowds out prediction, this is, I guess, what he would want to say.[20] This would suggest that disunification in the synchronic case involves more than just making a separate decision on each bet. It would also seem to involve some form of abstraction: While considering whether to accept a given bet, the agent abstracts from his decisions on the other bets on the table. (Indeed, without some such assumption it is difficult to see how an agent who violates standard probability axioms could be exploited in a synchronic setup to begin with.)

However, this potential difference between the diachronic and the synchronic case does not affect the issue of the availability of a dominating

[20] For his exposition and defense of that thesis, see Levi 1989, 1991, and 1997. For a critical discussion, see Joyce 2002 and Rabinowicz 2002.

option. The two setups can be analogous in this respect. In each setup, the dominating option can be available to the agent but is not an alternative he considers in his (disunified) deliberation. If this analogy is possible, then Levi has no grounds for his claim that the synchronic pragmatic arguments can establish the rationality of some constraints on agent's beliefs and preferences, but the diachronic arguments can't. Rather, it seems, the two kinds of arguments are on the same footing. As a matter of fact, as intimated in the introduction to this chapter, I don't think that pragmatic arguments of any variety are able to establish the inherent rationality of constraints. Instead, their proper function is to identify conditions that we have reason to comply with to afford being disunified in our decision making. In Levi's terminology, it is a matter of "tax avoidance." But I share his view that avoiding tax at all costs is unreasonable, especially if it is a case of a condition that it sometimes is reasonable to violate for noninstrumental reasons. (The principle of reflection is surely a case in point.) In this respect, however, synchronic and diachronic pragmatic arguments do not differ in any essential way, as their perspective on constraint violations is purely instrumental. But, to the extent that diachronic unification as a rule is less easy for us to achieve than the synchronic one, diachronic arguments provide stronger instrumental reasons for compliance.

ACKNOWLEDGMENTS

I am indebted to several people for comments and discussion. Luc Bovens and John Broome have been especially helpful in this respect. My work on this chapter was supported by a research grant from the Bank of Sweden's Tercentenary Foundation, and it was completed during my stay at the Swedish Collegium for Advanced Study in the Social Sciences (SCASSS) in Uppsala. I am indebted to the participants of my seminar at SCASSS in May 2004 for their useful comments. Some of the material in this chapter was presented earlier at the Ramsey Centennial conference in Cambridge in 2003 and at a workshop on philosophy and probability in connection with the GAP conference in Bielefeld that same year. I would like to thank the organizers of these events, Hugh Mellor and Luc Bovens with Stephan Hartmann, respectively.

REFERENCES

Aumann, R. 1995. "Backward Induction and Common Knowledge of Rationality." *Games and Economic Behavior* 18: 6–19.

Aumann, R. 1998. "A Note on the Centipede Game." *Games and Economic Behavior* 23: 97–105.

Bar-Hillel, M., and A. Margalit. 1988. "How Vicious Are Cycles of Intransitive Choice?" *Theory and Decision* 24: 119–45.

Bicchieri, C. 1989. "Backward Induction with Common Knowledge." *Proceedings of the Philosophy of Science Association* 2: 329–43.

Binmore, K. 1987. "Modelling Rational Players: Part 1." *Economics and Philosophy* 3: 179–213.

Broome, J., and W. Rabinowicz. 1999. "Backwards Induction in the Centipede Game." *Analysis* 59 : 237–42.

Christensen, D. 1991. "Clever Bookies and Coherent Beliefs." *Philosophical Review* 50: 229–47.

Davidson, D., J. C. C. McKinsey, and P. Suppes. 1955. "Outlines of a Formal Theory of Value, Part I" *Philosophy of Science* 22: 140–60.

Joyce, J. M. 2002. "Levi on Causal Decision Theory and the Possibility of Predicting One's Own Actions." *Philosophical Studies* 110: 69–102.

Levi, I. 1988. "The Demons of Decision." *The Monist* 70: 193–211.

Levi, I. 1989. "Rationality, Prediction, and Autonomous Choice." *Canadian Journal of Philosophy*, suppl. vol. 19: 339–63.

Levi, I. 1991. "Consequentialism and Sequential Choice." In M. Bacharach and S. Hurley (eds.), *Foundations of Decision Theory*, pp. 92–122. Oxford: Blackwell.

Levi, I. 1997. *The Covenant of Reason*. Cambridge: Cambridge University Press.

Levi, I. 2002. "Money Pumps and Diachronic Dutch Books." *Philosophy of Science*, suppl. to vol. 69: s235–47.

Maher, P. 1992. "Diachronic Rationality." *Philosophy of Science* 59: 120–41.

McClennen, E. 1990. *Rationality and Dynamic Choice*. Cambridge: Cambridge University Press.

Pettit, P., and R. Sugden. 1989. "The Backward Induction Paradox" *Journal of Philosophy* 86: 169–82.

Rabinowicz, W. 1995. "To Have One's Cake and Eat It, Too: Sequential Choice and Expected Utility Violations." *The Journal of Philosophy* 92: 586–620.

Rabinowicz, W. 1998. "Grappling with the Centipede." *Economics and Philosophy* 14: 95–125.

Rabinowicz, W. 2000. "Money Pump with Foresight." In Michael J. Almeida (ed.), *Imperceptible Harms and Benefits*. Dordrecht: Kluwer. pp. 123–54.

Rabinowicz, W. 2002. "Does Practical Deliberation Crowd Out Self-Prediction?" *Erkenntnis* 57: 91–122.

Raiffa, H. 1968. *Decision Analysis: Introductory Lectures on Choice under Uncertainty*. Reading, Mass.: Addison-Wesley.

Ramsey, F. P. 1990. "Truth and Probability" (1926). In D. H. Mellor (ed.), *F. P. Ramsey: Philosophical Papers*. Cambridge: Cambridge University Press.

Reny, P. 1988. "Rationality, Common Knowledge and the Theory of Games." Ph.D. dissertation, Princeton University.

Reny, P. 1989. "Common Knowledge and Games with Perfect Information." *Proceedings of the Philosophy of Science Association* 2: 363–9.

Schick, F. 1986. "Dutch Bookies and Money Pumps." *Journal of Philosophy* 83: 112–19.

Schumm, G. F. 1986. "Transitivity, Preference and Indifference." *Philosophical Studies* 53: 435–7.

Schwartz, T. 1993. *The Logic of Collective Choice*. New York: Columbia University Press.

Skyrms, B. 1980. "Higher Order Degrees of Belief." In D. H. Mellor (ed.), *Prospects for Pragmatism*. Cambridge: Cambridge University Press.

Skyrms, B. 1986. "A Mistake in Dynamic Coherence Arguments?" *Philosophy of Science* 60: 320–8.

Sobel, J. H. 1993. "Backward Induction Arguments in Finitely Iterated Prisoners' Dilemmas: A Paradox Regained." *Philosophy of Science* 60: 114–33.

Teller, P. 1973. "Conditionalization and Observation." *Synthese* 26: 218–38.

Van Fraassen, B. 1984. "Belief and the Will." *Journal of Philosophy* 81: 235–56.

20

Levi on the Reality of Dispositions

Johannes Persson

Isaac Levi is more interested in inquiry and how it progresses than he is in metaphysics. Questions concerning the role of disposition predicates in inquiry are more central to him than those concerning the nature and reality of dispositions. It has not stopped him from giving me and others very useful metaphysical advice. Currently, where empirical metaphysics is in vogue, there is every reason to see whether the two forms of philosophical interest might interlock substantially.

Levi has stimulating ideas indeed on the two forms of philosophical interest, and has recently summarized them in the slogan: "The reality of dispositions is a work in progress" (Levi 2003, p. 152). We can learn much about what kinds of dispositions are acceptable from tracing and comparing the histories of successful and less successful disposition predicates in scientific inquiry.

Levi explores one route along which dispositions become real. His idea is that the introduction of dispositions facilitates covering law explanation by increasing the number of laws. The successful disposition predicate eventually becomes integrated in scientific theory, much like an ordinary theoretical term, whereas the unsuccessful does not. My impression is that Levi thinks that this is the only way a disposition can become real. To evaluate this claim, an alternative course suggested by Jon Elster is introduced. I then try to bring out the differences between Levi's and Elster's views on dispositions, partly by suggesting that they resemble two aspects of full explanations discussed by Wesley Salmon. But more about that below.

1. LEVI ON STOPGAP EXPLANATIONS AND THEIR INTEGRATION

Introductions of dispositions in science are correlated with explanatory attempts. As Levi states, a disposition predicate is often introduced in order to cover an otherwise unexplainable gap. Take two species of fish vulnerable to predation by pike, such as tench and crucian carp. The crucian carp changes its body morphology in a special way when predators are around. It starts to grow "vertically" instead of "horizontally," that is, it becomes deeper bodied. The body-depth index changes so dramatically that earlier biologists thought the two morphs were different species. Within weeks the change is visible to the careful observer. But tench do not change in this way nor do most other fish we know of. Why? Since the two species exist in the same environments, a promising start is to assume that there is an intrinsic difference between crucian carp and tench. We assume that crucian carp have a special kind of built-in growth mechanism or disposition. In the presence of pike this mechanism is somehow activated and the fish becomes deeper bodied. A suitable disposition predicate matching this mechanism is the notion of a predator-induced "morphological defense," as in Brönmark and Miner (1992).

Let us focus on what kind of explanation the disposition predicate provides. Levi calls it "stopgap" explanation, and if I read him correctly this means both of the following two things. The stopgap explanation is preliminary, that is, by using it we commit ourselves to the idea that it eventually needs more work. The stopgap explanation is also of covering law type. One of Levi's examples concerns coin tossing. We know that some coins have the ability to land on edge when tossed. Therefore, whether or not all coins land heads or tails when tossed near the surface of the earth, the fact that a coin was tossed together with this regularity are not explanatory of the fact that it landed heads or tails. There is thus a gap in our explanatory powers, and in order to close it a disposition is introduced. We start to distinguish between coins with the surefire disposition to land heads or tails if tossed and other coins. Explanation is possible again by using the covering law:

$(x)(t)\{x$ has the surefire disposition to land heads or tails if tossed at t $\supset$ [x is tossed at $t \supset x$ lands heads or tails at $t(+\Delta t)$[1]]}

Besides being a covering law this formulation is a postulate that characterizes the dispositional predicate. Disregarding problems concerning determinism in ecology, a similar postulate can accompany the predator-induced "morphological defense":

$(x)(t)\{x$ has a predator-induced morphological defense at $t \supset [x$ is exposed to a pike at $t \supset x$ becomes deeper bodied at $t\,(+\Delta t)]\}$

[1] We should perhaps add this extra time now and in the parallel cases, although it is not included in Levi.

According to Levi, this constitutes a first, but only a first step needed for accepting the reality of the disposition.

(L1) Levi program in progress, step 1:
A disposition predicate providing a stopgap explanation is introduced.

To understand Levi's view it is instructive to compare the disposition predicate in (L1) with *ceteris paribus* clauses in laws and regularities such as "*ceteris paribus*, all mammals give birth to live young." The two have much the same function. Primarily, both fit the explanation into the covering law model of explanation: "They are place-holders for unspecified standing conditions in purely general laws," say Levi and Morgenbesser (1964/78, p. 400). But there are also differences. Even if *ceteris paribus* clauses occur in many and widely different laws, they are not assumed to be identical across contexts. Applying a disposition predicate in more laws than one involves a commitment, if revision takes place, to replace the disposition predicate by the same specification of standing conditions in each of the laws in which the disposition predicate appears (unless, of course, one loses faith in the predicate and comes to the conclusion that it was in fact disjunctive). Another notable difference is that while a *ceteris paribus* clause remains silent as to the location of the unspecified standing conditions, a disposition predicate locates these conditions in the object. That, at least, is granted by Levi and Morgenbesser: "[I]f we are told that fragile objects break when tapped lightly, we assume that if we are to improve or replace the generalization we should investigate the micro-structure of fragile objects" (ibid., pp. 401–2). This is certainly true of many good examples. By using the disposition predicate predator-induced "morphological defense," Brönmark and Miner accordingly locate some unspecified conditions to crucian carp, and why not to their micro-structure? Nothing similar is implied when we say that, *ceteris paribus*, all mammals give birth to live young. What must be kept equal here might be external factors.

Let us now continue with the subsequent steps of the Levi program, for it is clear that we cannot rest content with having taken the first step. That kind of disposition will not do in the longer run. "In that setting, explanation by disposition becomes suspect in the way that explaining the responses of those who imbibe opium by appealing to its dormitive virtue is" (Levi 2003, p. 139).

Each point of inquiry from L1 and onward belongs to one of three stages with respect to what more we claim to have filled in about the disposition. Since this is easiest to understand in connection with what is often called the basis of the disposition, I first follow Levi and Morgenbesser (1964/78,

p. 403) in distinguishing between, on the one hand, the innocent or *problem-raising* situation after the first step (L1) where no basis is known to exist and no claims that there is one have been made and, on the other, later *problem-solving* situations where legitimate bases have been found. In between lies Ryle's *mystery-raising* stages where no basis is known to exist, although claims that there is one are being made. Problem-solving stages are mystery-solving, too, so the idea of progress we are looking for can be pictured as a development from problem-raising or mystery-raising to problem-solving and mystery-solving.

In Levi (2003) the micro-basis idea is downplayed. He now claims more generally that giving up one's expectation of progress after L1, and saying that this is the best we can achieve, transforms something that was initially problem-raising to the level of being mystery-raising. Mystery raising occurs in all situations where inquirers have not integrated the disposition predicates into an explanatorily adequate theory, and yet they judge the explanations they offer as satisfactory. (Levi's most important criticism of Elster's alternative view below is that it, in effect, amounts to certifying mystery raising in this sense.)

The change in Levi's thoughts on this matter is probably to the better. But the concept of "integration" is more difficult to come to grips with than the earlier story about micro-bases. To begin with, it presupposes an idea about what kind of theory the disposition predicate is to be integrated into. Levi has little more to say about this than that the theory should be part of the established body of full belief.

(L final) Levi program in progress, final step:
The disposition predicate is adequately integrated into a theory that is part of the established body of full belief.

This admits of a pluralistic view of science. The nature of integration can take many forms depending both on the structure of the theory and on the specific demands and goals of the research program (cf. Dupré 1993, pp. 261–4).

To analyze the degree of integration in actual cases is not a simple matter. However, returning to the crucian carp, it is evident that many attempts to integrate the disposition have been made. This work has partly taken the form of explaining why the emergence of such a defense is fitness-increasing. How efficient it is in avoiding predation is a central piece of knowledge. And the effects of prey size on pike behavior and how this relates to optimal foraging models have been examined in this research program, for instance, in Nilsson and Brönmark (1999). Why it is not a constitutive defense has been answered by investigating the costs of becoming deeper bodied, how predation pressure varies, and so on. The kinds of cues that trigger the disposition have also been investigated. It was shown in Brönmark and Petersson (1994) that the cues are chemical rather than visual, and it was furthermore confirmed that the cues are released not by injured conspecifics or from the predator per se, but only from predators with a piscivorous diet. This makes the disposition adaptive in habitats with a suite of predators or when predators undergo niche shifts. For instance, if the perch and pike in a pond feed on, say, crayfish and other invertebrates, the morphological defense is not triggered. The relevant cue is not released until the predators start feeding on fish, that is, when the individuals become piscivores.

2. DISPOSITIONS AS WAYS IN WHICH THINGS HAPPEN

Contrary to Levi, some have assumed that explanation by mechanism or disposition is a *substitute* for covering law explanation. Jon Elster is one example. While Elster employs the concept of mechanism that I am sympathetic to, I believe that what he says can be translated roughly into the language of dispositions. Levi (2003) makes the same point in his discussion of Elster's theory.[2]

First, an introduction to Elster's view. The background against which Elster proposes his account is captured in his answers to the two questions: Are there lawlike generalizations in the social sciences? If not, are we thrown back on mere description and narrative? Elster's answer to both is negative. The ideal of lawlike explanation in history and the social sciences is implausible, claims Elster. To escape explanatory nihilism the idea of a mechanism is introduced. It is supposed to do its explanatory work at a level between law and description (Elster 1999, p. 1).

The relation between laws and dispositions is, as we have seen, a very close one for Levi, but Elster disconnects them. The perhaps clearest formulation is from his 1989 book:

> Laws by their nature are general and do not suffer exceptions. One cannot have a law to the effect that "if *p*, then sometimes *q*." Mechanisms, by contrast, make no claim to generality. When we have identified a mechanism whereby p leads to q, knowledge has progressed because we have added a new item to our repertoire of ways in which things happen. (Elster 1989, pp. 9–10)

[2] It is an interesting thought, but one that I cannot pursue at this point, that the translation is less than perfect. Evidence, but not an argument, for this is the fact that George Molnar, who was as liberal concerning interventions as Elster is, switched from "dispositions" to "powers" in later writings (Molnar 2003). Maybe mechanisms are more fundamental than dispositions. Mechanisms might give rise to dispositions but not vice versa. If Levi would agree, reconciliation between the two programs would be substantially facilitated.

In Elster's world, lawlike generalizations are scarce, while mechanisms abound. Moreover, his mechanisms often come in *pairs*. Some people prefer what they can have; others tend to want what they do not or cannot have. Here we seem to have one adaptive and one counteradaptive preference, nicely mirrored in the proverbs referring to sour grapes and forbidden fruit (Elster 1999, p. 7). Unreachable grapes are sour. Forbidden fruit is sweet. It is not a unique case. For any proverb one seems to be able to find one that asserts the opposite.

Elster's rationale for claiming that introduction of mechanisms does not add to the number of laws must be that mechanisms can interact in many ways. Lawlike generalizations are scarce partly because mechanisms abound. There are two characteristic ways in which mechanisms may intervene: type A and type B. Type A interventions are preemptive. If the disposition to have adaptive preferences is triggered, the disposition to have counteradaptive preferences is not (and vice versa). Type B interventions are modifiers. The net effect differs from the effects of the component dispositions. There is a variety of type B interventions. Elster (1999, p. 8) distinguishes between B_1 mechanisms, where one (external) cause triggers both component dispositions, and B_2 mechanisms, where one component triggers the other. Both type A and type B mechanisms make prediction problematic. In the first case the indeterminacy concerns which (if any) of several causal chains will be triggered. In the second case we can predict the triggering of two causal chains, but the net effect is indeterminate. In the aquatic example, the demands to forage and to avoid predation govern the behavior of fish. But while the triggering of these dispositions is easy to predict, the net effect is not.

Elster's view is attractive too. Prima facie, there is at least one other reason why dispositions are introduced besides providing stopgap covering law explanations, namely, to explain how things happen. This alternative origin of dispositions should also be kept track of.

(E1) Elster program in progress, step 1:

A disposition predicate identifying a way in which things happen is introduced.

3. TWO DIFFERENCES BETWEEN LEVI AND ELSTER

There are real differences between Levi and Elster. Most obviously, Levi's dispositions are surefire, while Elster's mechanisms are not. The latter can be combined, and type A and B mechanisms cannot be surefire nor can their components. A consequence of Levi's committal to surefire dispositions is that he cannot accept configurations of A and B type. Levi's position is therefore exposed to many of the counterexamples currently in vogue. If neither A nor B interventions are admissible, then the existence of *finks*, *antidotes*, and *intrinsic maskers* must also be denied.

A finkish disposition ceases to exist on the instantiation of the disposition's characteristic stimulus condition, and so the manifestation or response is not brought about. C. B. Martin's (1994) electro-fink example is paradigmatic. Here is a version of it: We have an electrical wire disposed to transmit a certain wattage for a certain duration. To this wire we connect a sensitive and fast-acting circuit-breaker (the fink). Now, the stimulus (=closing the circuit) that would normally trigger the disposition's manifestation (=transmitting the wattage) also triggers the fink (=the circuit-breaker), which makes the disposition to transmit the wattage disappear, and the manifestation process is canceled. It takes some time before the fink reacts, but it also takes some time to transmit the wattage, and it can be aborted before completion (cf. Lewis 1997).

While not put to the test, most of us would say that the fink and the disposition can exist side by side as two dispositions had by different objects. The fink does not influence the disposition of the wire at that time. But conditional accounts of Levi's kind would not say that. I agree with Levi (2003, p. 149) that coming to know that there is an electro-fink makes one disposed to abandon the claim that the wire has the surefire disposition to transmit a certain wattage for a certain time. But whereas he thinks that what is abandoned is the belief that the wire has a disposition, I think that what is abandoned is the belief that this disposition is surefire.

Similarly for antidotes. An antidote interferes with the operation of the disposition without destroying its causal basis. It merely breaks the causal chain leading from stimulus to response (Bird 1998). An antidote to a poison may change the patient's physiology so the poison cannot do the damage it normally does or by repairing the damage done before it can result in illness or death. Intrinsic maskers, finally, can neither be admitted by Levi. George Molnar (1999, p. 5) scans Greek mythology and comes up with Tantalus, whose ability to drink was masked by his disposition to cause all fluids he approached to evaporate; and King Midas, whose disposition to turn everything he touched into gold masked his ability to nourish himself. In a later book, Molnar (2003, p. 93, footnote) even adds the case of an intrinsic masking of an intrinsic masker:

> Among the satellites now in orbit is one designed to receive infrared signals from deep space. As the satellite's own heat system would mask the incoming signals, the antennae by which the signals are received have to be continuously cooled with liquid nitrogen. This cooling is a case of intrinsic masking of an intrinsic masker.

Finks, antidotes, and maskers are not acceptable to Levi, at least not in an early stage of inquiry. But they all fit nicely with Elster's view. Plainly, all three are mechanisms of type A (the philosophy of dispositions could have use for some of his type B mechanisms as well). Thus, Levi's and Elster's approaches to dispositions clearly differ in this way.

The second difference between Levi and Elster is that Levi's dispositions add to the number of laws, while Elster's mechanisms identify ways in which things happen. Explicitly, Elster only claims that identifying a way in which things happen is to do more than just to describe but less than to state a law: It is to identify an easily recognizable causal pattern. As is seen below in the section on Salmon, I think the conception of dispositions as ways in which things happen promises more, but let us settle with what Elster explicitly claims for now. It is enough for creating a difference between the two views. Does it have further consequences?

Even if I haven't seen Elster making this point, dispositions as ways in which things happen seem sometimes to be introduced into research programs where there is a covering law. Sometimes the laws are felt to be of the wrong kind. Early research about crucian carp is a good instance in case. Because of the morphological differences, the two morphs were at that time classified as distinct species.

Already in 1838, a Swedish priest, C. O. Ekström, rejected the idea that there were two species of this carp. Not observing the actual change in particular individuals, he examined lakes where a transition from one type of fish to the other had taken place. Rather than concluding that one species had miraculously transformed into another, he conjectured that the different types of fish represented two morphs of the same species. Ekström (1839, p. 225) had a clear opinion on the more precise explanation, namely, that differences in resource levels affected the body morphologies of crucian carp.[3] This leads to the following two generalizations (ranging over crucian carps):

(CC1) $(x)(t)\{$the habitat of x is resourceful at $t \supset x$ develops into the deeper bodied morph at $t(+\Delta t)\}$
(CC2) $(x)(t)\{$the habitat of x is not resourceful at $t \supset x$ develops into the streamlined morph at $t(+\Delta t)\}$

One interesting thing about CC1 and CC2 is that they seem to be true, at least in natural environments. It is even the case that some contemporary

[3] "Att Dam-rudan, försatt i frihet, återtager sin naturliga breda form, synes äfven deraf, att ju oftare hon omplanteras i nya dammar, der öfverflöd på näring finnes, desto bredare blir kroppsformen."

Finnish ecologists still hold them to be explanatory of the fact that there are two different morphs of crucian carp. And these covering laws may well be true also on the morphological defense story. Northern pike are very efficient both as colonizers and as predators, but they are gape-limited so there is a size refuge for prey. In Scandinavia you will hardly find a lake without pike. Crucian carp have been of economic interest. Farmers used to put them in all kinds of more or less temporary ponds with comparatively low resource levels, especially after some time, since the populations were normally dense. Since crucian carp is extremely vulnerable to predation, we will expect to find only two kinds of environments where they exist: On the one hand, in pikeless ponds where they form dense populations; on the other, in lakes with pike and comparatively few individual crucian carp. Resources are sparse in the first kind of habitat and rich in the other. Ekström even called the two morphs the pond and the lake morph, and if we disregard the morph/species distinction, "Dam-ruda" and "Sjö-ruda" were already established Swedish names for them.

It is implausible to think that Brönmark and Miner introduced their disposition in an effort to provide a stopgap explanation along the lines suggested by Levi's L1: No relevant gap seems to have existed. It is more likely that E1 motivated the introduction: They wanted to find a way – or maybe the way – in which this morphological change happens.[4]

4. ETIOLOGICAL AND CONSTITUTIVE ASPECTS OF EXPLANATIONS

Wesley Salmon (1984, pp. 269–70) usefully distinguished between "etiological" and "constitutive" explanations. Etiological explanations explain a given fact by showing how it came to be as a result of antecedent events, processes, and conditions. A constitutive explanation, on the other hand, does not explain in terms of antecedents. The explanation shows, instead, that the fact to be explained is constituted by underlying mechanisms. For instance, according to Salmon, many cases of physical reduction qualify as constitutive explanations. Both etiological and constitutive explanations are relevant aspects of the full explanation. For illustration, let us have a look at Salmon's (1984, pp. 270–1) own example.

To give a full explanation of the destruction of Hiroshima near the end of World War II, it would be necessary to refer to an atomic bomb and to explain

[4] From what I have learned from professional ecologists, it is not uncommon to suggest to a doctoral student that he or she should first establish a good and interesting correlation in the area of interest, and then go on to suggest mechanisms explaining it.

the explosion in terms of the assembly of a critical mass of U-235. Such an explanation would embody constitutive aspects. The explosion is explained in terms of a self-sustaining chain reaction, and this notion is causally explained in terms of the mechanisms of nuclear fission. The same explanation of the destruction of Hiroshima would include reference to the dropping of the bomb from an aeroplane and the detonation by implosion of a critical mass of fissionable material at a certain place above the city. These are etiological factors because they are antecedent events that contributed causally to the occurrence of the explanandum-event.

From this point of view, whether one opts for L1 or E1 has perhaps to do with the kind of explanatory project or explanatory phase one is involved in. It is to be expected for historic reasons that covering laws occur in etiological explanation and that L1 marks an etiological interest. L1 is deep-rooted in a Humean tradition. In Elster (1983) it is also clear that E1 implies work on the constitutive aspects of explanation. The search for mechanisms in the social sciences was connected to the program of methodological individualism – the idea that all social phenomena can be explained in terms of individuals and their behavior. Nowadays, Elster is not equally outspoken about reductionistic matters, but it is still fair to say that E1 is connected to constitutive explanation. To the extent that Elster is moving away from that conception, there are others interested in pursuing projects similar to the original. John Dupré (1993, p. 106) is one example:

> Reductive explanation is required to account for *how* things of a certain kind do what they do; but they typically do not help us to understand or to predict *what*, among the behaviours of which it is capable, a complex thing will do.

What L1 and E1 primarily teach us, according to this view, inspired by Salmon, is that disposition predicates can be used in both etiological and constitutive explanation. Can E1 and L1 also be understood as two starting points leading to the same kind of complete explanation, that is, is the idea of a full explanation applicable? It is to be noted that Levi accepts the possibility of intervening dispositions, such as finks, at later stages when but only when the original stopgap explanation can be removed without cost. At L final the disposition predicate has lost its placeholding function entirely. When this happens, opposing or masking dispositions provide no further threat to the inquiry. So for Levi the bridge offered is okay so long as L1 comes first. Elster should also be in favor of reconciliation, although he is skeptical about the possibility of finding lawlike generalizations in the special sciences: "[M]echanisms are good only because they enable us to explain when generalizations break down. They aren't desirable in themselves, only *faute de mieux*"

(Elster 1999, p. 6). It seems then that E1 and L1 could be understood as two starting points ideally leading to the same full explanation.

Here is a possible objection to that project. The explanans of the constitutive explanation in Salmon's example concerns the assembly of a critical mass of U-235. What explanandum does it fit? Arguably, it primarily fits an explanation of the bomb's property of being explosive. The self-sustaining chain reaction, on the other hand, is part of the explanans of an event, the explosion of the bomb, which is a different explanandum. A distinction in terms of different explananda is there to make. On the one hand, we seem to explain either a property or an object by explaining how it is constituted; on the other hand, we explain an event or change etiologically by explaining why it took place. Many examples of this kind of distinction and its consequences can be found in Dupré's 1993 book. But he differs from Salmon in being skeptical about the possibilities for constitutive explanation to function in an etiological explanation. Recall the last part of the above quotation: "[B]ut they typically do not help us to understand or to predict what, among the behaviours of which it is capable, a complex thing will do." Dupré continues: "The latter is generally to be addressed in terms of the autonomous understanding of the phenomena at the higher level" (Dupré 1993, p. 106).

If the objection is correct, etiological and constitutive explanation are on different tracks. There is no bridge between them. In one sense the problem concerning the differences between Levi's and Elster's views of dispositions then dissolves. There is no need to reconcile the two perspectives, since they will never meet halfway. Salmon's ideas have been useful but in a negative way. Levi examines dispositions functioning on the etiological level. Elster and/or Dupré propose an account of dispositions on the constitutive level. But that result would be troublesome for later phases of Levi dispositions becoming real. How can integration of initially problem-raising disposition predicates be achieved if we assume that we have an autonomous understanding of the phenomena occurring as stimulus and response? Levi should be careful not to follow Dupré's lead.

And Dupré is entitled only to the claim that we cannot *in general* assume that constitutive and etiological explanations can be combined. There may still be plenty of cases where Salmon's ideas are applicable. In those cases it provides a bridge from which potentially both Levi and Elster might benefit.

It is almost a dilemma. Dupré's view presents one horn and Salmon's view can be seen as the other. Besides offering a bridge that is useful *eventually*, the notion of full explanation accentuates the tension between L1 and intervening dispositions *right from the start*. I am not sure that Levi can live happily with the idea that inquirers should stick to the idea that dispositions generate covering laws in such circumstances. I think that in a step closely following L1, that is, much earlier than L final, Levi should be prepared to make suitable changes. If this can be done at limited cost, things look bright. If not, then maybe Levi should give up L1 altogether.

That would not be the end. There would still be an interesting story to tell about integration. The reality of dispositions could be another kind of work in progress leading to the same outcome. To say with Levi that the reality of dispositions is a work in progress is to say that dispositions are real only when their placeholding mission has been accomplished (Levi 2003, p. 152). Speaking metaphysically, Levi is in favor of type A and type B. Only real dispositions can handle such cases because real dispositions are not restricted by L1.

5. DISPOSITIONS: METAPHYSICS AND INQUIRY

Everything but the previous paragraph seems to have been driven by epistemological considerations. Does Levi's approach tell us anything about the *metaphysics* of dispositions? That depends on what we include in the metaphysical study. My impression is that we include much more than is usually acknowledged. To begin with, many discussions in metaphysical texts concern *existence criteria*. For one example, versions of the Eleatic Stranger's test (the mark of being is power) abound in contemporary metaphysics and especially in the metaphysics of dispositions. Another illustration is found in work on Ramsey sentences. Some metaphysicians take these as providing existence criteria for what properties there are. To answer that question, D. H. Mellor takes all predicates in statements of laws of nature. He then conjoins them, replaces all the predicates with variables, and gets a Ramsey sentence that says that "there are in the world properties that occur in this and that way in laws of nature." "So, for me," continues Mellor in a *Theoria* interview, "this Ramsey-sentence provides an existence criterion, i.e. a claim about what determines what factual properties there are in the world. I think we need such a criterion, because without one it's too easy to postulate properties without having any clear idea of what counts as a property, or what determines whether some property you've postulated really exists" (Maurin and Persson 2001). In the present context, Levi (2003, p. 141) discusses the possibility that successful integration of disposition predicates gives rise to two characteristics also figuring as well-known criteria for properties: The problem-solving steps make safe that dispositions display themselves in more ways than one and that two objects that differ with respect to one property must differ with respect to another. Questions concerning existence criteria seem as epistemological as they are metaphysical in the sense that they use fundamentally epistemological considerations to motivate metaphysical conclusions. There is nothing particularly strange about this. We also use explanatory considerations, arguments from simplicity, and so on, in metaphysical enterprises. There is, to my knowledge, no exclusive metaphysical machinery. Moreover, students of metaphysics do not employ more rigorous acceptability criteria. They cannot, since good evidence is so hard to come by in metaphysics.

A more promising suggestion would be that metaphysical *questions* typically differ from other kinds of questions in philosophy and science. "What makes the descriptions given in a scientific theory true?" is unlikely to be found in fields of inquiry other than metaphysics. For one thing, it is more general than most similar questions posed within that science.[5] According to this suggestion, whether Levi's approach tells us anything about the metaphysics of dispositions depends on whether the questions he poses could appear in metaphysical inquiry (or whether the kind of conclusions he comes to could function as answers to questions posed in metaphysical inquiry). It cannot be denied that Levi's questions more often than not are of the right kind. On the plausible assumption that the theories we integrate into are supposed to be true, Levi's account interlocks nicely with metaphysical interests.

CONCLUSION

More than thirty years ago Isaac Levi made an intriguing suggestion about how further inquiry may improve on a problem-raising disposition predicate's chances of eventually being taken to express a real property. First, it should function in a stopgap explanation. For a while, as long as nothing better is suggested, it will be safe in that position. But this is because it also carries a promise of transforming into a problem-solving predicate in the future. Should this promise not be fulfilled, or should the participants in the program suddenly decide that it already constitutes a satisfactory explanation, it degenerates to the point that it becomes mystery-raising. By being adequately integrated, on the other hand, we have every reason to think that the disposition is real, and this is what Levi intends with the slogan: "The reality of dispositions is a work in progress."

Levi does not, however, provide the complete story of the rise and fall of disposition predicates. Especially the part about how disposition predicates are introduced does not cover all the cases captured by Elster's suggestive remarks about ways in things are. As it stands it is a conception of dispositions

[5] Compare Mellor in the *Theoria* interview.

that seems to be in open conflict with Levi's postulate in L1, which, as we have seen, is problematic also for other reasons. I think of Elster's suggestion as an important addition to Levi's picture and that a weakness of Levi's current position is that it cannot allow for many interactions between different dispositions until the program is fully completed. I also think that it would be congenial to Levi's general approach to inquiry to develop this view of his on dispositions further so that several beginnings can be accepted.

REFERENCES

Bird, Alexander. 1998. "Dispositions and Antidotes." *Philosophical Quarterly* 48: 227–35.

Brönmark, Christer, and Jeffrey C. Miner. 1992. "Predator-induced Phenotypical Change in Body Morphology in Crucian Carp." *Science* 258: 1348–50.

Brönmark, Christer, and Lars B. Petersson. 1994. "Chemical Cues from Piscivores Induce a Change in Morphology in Crucian Carp." *OIKOS* 70: 396–402.

Dupré, John. 1993. *The Disorder of Things*. Cambridge, Mass.: Harvard University Press.

Ekström, C. O. 1839. "Iakttagelser öfver formförändringen hos Rudan (*Cypr. Carassius* Lin.)." In *Kongl. Vetenskaps-academiens handlingar för år 1838*, pp. 213–25. Stockholm: Norstedt & Söner.

Elster, Jon. 1983. *Explaining Technical Change*. Cambridge: Cambridge University Press.

Elster, Jon. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Elster, Jon. 1999. *Alchemies of the Mind*. Cambridge: Cambridge University Press,

Levi, Isaac. 2003. "Dispositions and Conditionals." In H. Lillehammar and G. Rodriguez-Pereyra (eds.), *Real Metaphysics*, pp. 137–53. London: Routledge.

Levi, Isaac, and Sydney Morgenbesser. 1964/78. "Belief and Disposition." In Raimo Tuomela (ed.), *Dispositions*, Synthese library no. 113, pp. 389–410. Boston: D. Reidel.

Lewis, David. 1997. "Finkish Dispositions." *Philosophical Quarterly* 47: 143–58.

Martin, C. B. 1994. "Dispositions and Conditionals." *Philosophical Quarterly* 44: 1–8.

Maurin, Anna-Sofia and J. Persson. 2001. "Realistic Metaphysics: An Interview with D. H. Mellor." *Theoria* 67: 96–113.

Mellor, D. H. 2003. "Real Metaphysics: Replies." In Lillehammar and Rodriguez-Pereyra (eds.), *Real Metaphysics*, pp. 212–38. London: Routledge.

Molnar, George. 1999. "Are Dispositions Reducible?" *Philosophical Quarterly* 49: 1–17.

Molnar, George. 2003. *Powers: A Study in Metaphysics*, ed. Stephen Mumford. Oxford: Oxford University Press.

Nilsson, P. Anders, and Christer Brönmark. 1999. "Foraging among Cannibals and Kleptoparasites: Effects of Prey Size on Pike Behaviour." *Behavioural Ecology* 10: 557–66.

Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

21

Replies

Isaac Levi

PREFACE TO MY REPLIES

I wish to express my appreciation to all the contributors to this volume for their contributions. The quality of the chapters and the distinction of the authors calls for more extended comment on what they have to say than the space allotted to me permits, and this may sometimes lead to my stating my reactions with a bluntness that disguises the gratitude, regard, and respect that I have for all of the authors. I am aware that my views are often controversial, and the authors have as a group risen admirably to the occasion by fueling the controversy.

WHAT SORT OF REALIST IS PEIRCE AFTER ALL? REPLY TO MISAK

Cheryl Misak and I share an admiration of Peirce's account of fixing belief – an account whose basic elements remained constant throughout the bulk of Peirce's career. According to that account, changing beliefs requires justification. Current beliefs do not need justification. Indeed, rational inquirers judge all their current beliefs to be true. This shared appreciation is qualified by substantial disagreement concerning how to understand Peirce's elaboration of his ideas – disagreement that is of some philosophical moment.

In "The Fixation of Belief," Peirce wrote that "the sole object of inquiry is the settlement of opinion." He explicitly dismissed the thesis that the aim was to settle opinion with true (i.e., error-free) belief. He did so in the context of a comparison of four different methods of relieving doubt. Three of them do not take the truth-value of a candidate belief into account in determining whether it affords legitimate relief. The scientific method does. It self-consciously strives to correct its own errors. When difficulties arise that legitimately provoke doubts about conclusions obtained through scientific inquiry, Peirce maintained that following scientific methods guarantees that either the conclusion challenged is true or further testing will with probability detect the error. In spite of his allegation that in inquiry not true belief but only belief (i.e., relief from doubt) is sought, he clearly and explicitly declared in the very same essay allegiance to the scientific method that corrects its own errors according to procedures that he sought to explain in subsequent essays to which "The Fixation of Belief" is an introduction. His dismissal of the thesis that the aim of inquiry is to settle opinion with true belief is not inconsistent with this. He sought to *argue* for the superiority of the scientific method over alternatives and wished to avoid begging the question by presupposing the propriety of following the aims of the scientific method.

What does true belief amount to for Peirce? Misak and I agree that Peirce did *not* say that asserting that h is true is equivalent to claiming that if inquiry according to scientific method were to proceed indefinitely and converge to a limit, h would be in the theory at the limit. He insisted that such hypothetical convergencc would be at best a "hope," and, in any case, the convergence would be with probability 1 in the sense of the strong law of large numbers – that is, almost certainty. These explicit acknowledgments by Peirce preclude such definition. This did not prevent Peirce from regarding the ultimate goal of inquiry to be the acquisition of a true, complete theory at the End of Days. He explicitly embraced such an ideal of progress and heaped encomia on the altruistic pursuers of the ideal in the name of the community of inquirers. Peirce did not define truth as the final opinion. But he thought that progress toward such a Messianic opinion is a mark of scientific inquiry. Misak disagrees not only with Messianic Realism but with the claim that Peirce was such a Realist.

Peirce's conception of truth as I reconstruct it (and I hasten to emphasize that it is a reconstruction) maintains that an inquirer X's theory of truth is a Tarski-like theory where X's current full beliefs are taken to be part of the truth theory. Since X's belief state is subject to change, so does the truth theory. As Quine put it, truth is judged relative to the evolving doctrine.

I claim that it is in this sense that Peirce thought that "truth" and "falsity" are definable in terms of doubt and belief. In this sense, "definable" does not, of course, mean "reducible." Peirce might also have thought that specifying the way in which truth and falsity figure in characterizing the aims of inquiry render "truth" and "falsity" definable in terms of doubt and belief. Peirce wanted the alethic notions to be understood in a manner that renders them relevant to the justification of changes in belief.

I think that Misak's interpretation of Peirce's conception of "true belief" as a "catch-all" in the sense that "were a belief or theory to satisfy all our local aims in inquiry, it would be true" is extremely uncharitable to Peirce.

Since the "local aims in inquiry" when taken over time often are in conflict, they cannot all be satisfied. There is a way out. One might say that all local aims in inquiry are directed toward true belief. The position then becomes trivial.

Misak also claims that that when we aim at truth we aim at beliefs that would be forever stable – that is, doubt-resistant. But doubt-resistant beliefs are incorrigible. So according to Misak, Peirce advocated seeking incorrigible beliefs. The best way to do this is to restrict belief to logical truth and any conceptual necessities that are firmly entrenched. In any case, Misak herself insists that according to Peirce, scientific methods of fixing belief are full of surprises. They lead to conflict and instability just as surely as other methods do.

I agree with Misak that nothing in logic required Peirce, who took a common feature of proximate aims of efforts to relieve doubt according to scientific method to be to replace doubt by true belief, also to insist that scientific inquiry embraced as an ultimate goal progress toward the true complete story at the End of Days. It is simply a fact expressed over and over again by Peirce that he thought that scientific inquiry was progressive in the sense that it has progressive aims and is based on the hope that these aims can be fulfilled in the limit. Scientific inquiry seeks progress in approaching the true, complete state T of belief. That is the ultimate goal of inquiry.

Peirce's Messianic Realism is in conflict with his view that X's beliefs are corrigible. I summarized the argument in my comments on Fuhrmann's discussion of James. The importance of emphasizing the conflict is that there were and continue to be many authors (the best known in the post–World War II era is Karl Popper) who embrace some kind of progressive Messianic Realism without appreciating the problems it poses.

Dewey was not a Messianic Realist and, indeed, did not make truth an important feature of even proximate aims of local inquiries. Dewey does not appear to have been even a Secular Realist. André Fuhrmann thinks that James, like Peirce, was a Messianic Realist. According to Fuhrmann, James abandoned Peirce's commitment to the use of current beliefs as a standard for judging possibility when contemplating belief change in the long run. If Fuhrmann is right, James betrayed the belief-doubt model for inquiry. If Misak is right about Peirce (counter to fact, so I think), Peirce was closer to Dewey than I think can be sustained by the literary evidence.

I prefer Secular Realism to the antirealism of Dewey and of Misak's Peirce. Secular Realism is congenial with Peirce's belief-doubt model and with his claim that scientific methods are self-correcting and with an understanding of inquiry as seeking to avoid false belief that respects Peirce's stipulation that a conception of truth should be relevant to belief and doubt without crude reduction to belief or satisfaction of our aims. Only Peirce's idea of progress has to go.

At one time I thought that the passages focusing on "vital matters" from 1898 were evidence that Peirce recognized the tension among corrigibilism, the belief-doubt model, and Messianic Realism. I have become convinced that Misak is right to discount this textual evidence. Peirce wrote his remarks in response to James's "Will to Believe." The vital matters he discussed related primarily to theological questions where suspension of judgment is not an option. He seems to have thought that such vital questions are considered so in the heat of the moment when passions are at some peak. James's arguments appear less compelling when the agent is able to come down and reason in a cool fashion. Understood in that fashion, I think that Peirce's remarks do not amount to a betrayal of the belief-doubt model. Thus, I concede to Misak that Peirce did not appear to recognize the tension between his Messianic Realism and his belief-doubt model. Nonetheless, Peirce's advocacy of both was incoherent.

Misak contends that although Peirce and I "share the anti-Cartesian view that we fully believe what we do not doubt," I "struggle" with the fact that agents often exchange full belief for something incompatible with it. Peirce, on the other hand, focuses on the kinds of things that can upset or unfix full belief.

Peirce's aperçus about surprise cannot disguise the fact that Peirce wrote much more about how to eliminate doubt than how to justify coming to doubt. The many philosophers and logicians who have worked on the topic of belief revision for the past third of a century have done more to elaborate on Peirce's insights concerning the "undoing" belief than Peirce ever contemplated. In the *Enterprise of Knowledge* I made a few suggestions about undoing belief. But it is only in response to the work of Alchourrón, Gärdenfors, and Makinson and the community of scholars touched by their ideas that my own efforts to work on how to give it up began in earnest.

I received my philosophical training during a period when it was widely and wrongly thought that pragmatists confused issues that analytic philosophy clarified. In the past thirty years, students of belief revision have addressed issues concerning change of view neglected by analytic philosophers. Contemporary pragmatists ought not to congratulate themselves on belonging to a tradition that anticipated the importance of these problems. They ought to join with those who actively pursue their solution.

Fuhrmann's discussion of James's conception of truth is an excellent elaboration of how James might have tried to respond to the objection I raised against Messianic Realism. Messianic Realism is the thesis that the ultimate aim of inquiry ought to be convergence on the true, complete theory of the world. Peirce was a Messianic Realist. Fuhrmann takes James to have been one. My objection to Messianic Realism is that in order to promote this ultimate goal, an inquirer should avoid giving up any full belief and, to the extent that efforts to expand the body of full belief has as a consequence the need to give up other full beliefs, the inquirer should avoid expansion as well. Messianic Realism, therefore, conflicts with the revisability of the inquirer's state of full belief – a thesis that Peirce, James, and Dewey all endorsed.

Two assumptions are salient in my argument:

1. When X is deliberating whether and how to change X's state of full belief, X does so on the assumption that X's current state of full belief contains no error and that X judges it impossible that X's current state contains error.
2. X is concerned to promote progress toward the true theory T – a theory that X is certain when X's belief state is K contains *all* the consequences of K. So incurring a risk of converging on a theory that entails $\sim p$ where p is a consequence of K is risking failure in achieving this ultimate goal – a risk for which there is no compensation.

Let inquirer X seek progress toward reaching the true theory T (whatever that may be). X's current state of full belief, certainty, or knowledge is K. Let X contemplate removing p from K where K has p as a consequence. According to assumption 1, there is no serious possibility that p is false regardless of whether p is a logical truth or an extralogical proposition.

If p is a logical truth, X anticipates from X's initial point of view that after giving up p, X's next state of belief will become incoherent. If p is extralogical, coherence may be retained by giving up p but $\sim p$ will become a serious possibility. If X were to stop there, X would avoid coming to believe anything false (as judged from X's initial point of view relative to K). But X would lose information without compensation for the loss. If X's ultimate aim is promoting progress toward reaching the true theory T as 2 says, X should attempt to expand the contraction $K - p$ of K. From X's point of view, when K is X's state of full belief, there is a serious possibility that X will import information entailing $\sim p$. That is to say, X will import information that X is certain is false. Worse yet, in subsequent investigation there is a serious possibility that X will never be extricated from this false belief. It does not matter that there is a serious possibility that X will escape the clutches of such error. If *both* assumptions 1 and 2 are in place, when X contemplates contracting by removing p from K, X is committed to regarding the risk of failing to promote progess toward the true doctrine to be sufficiently great to warrant refusal to contract – no matter how small that risk is.

To be sure, X might be compelled to remove p because information yielded by observation reports that $\sim p$. But according to assumption 1, making observations that mandate adding such information must run counter to progress toward the true complete story. One should avoid making such observations.

Thus, not only should belief systems not be revised, but they should never grow. On this latter point, Fuhrmann misunderstands me.

Fuhrmann is concerned to argue that Messianic Realism (assumption 2) can be consistent with the modifiability of states of full belief. However, he concedes that if we adopt assumption 1, "the issue as to whether p must be regarded as closed. But when contracting with a view to truth in the long run, we need reassurance that the case as to whether p, may be reopened." Fuhrmann argues for giving up assumption 1 and suggests that James does so. And Fuhrmann may be right about James's view.

My point is that giving up assumption 1 flouts the belief-doubt model. The belief-doubt model to which, Fuhrmann agrees, Peirce and James subscribe *requires* taking assumption 1 very earnestly.

Fuhrmann rebuts my argument by replacing 1 with a double standard for serious possibility. Fuhrmann thinks that while deliberating as to whether to remove p from K, one can judge the falsity of p to be a possibility. Fuhrmann acknowledges that it is not a serious possibility according to K. He suggests that in deliberating about ultimate goals, judgments of possibility should be judgments of "metaphysical" possibility by which Fuhrmann means judgments of serious possibility relative to the "urcorpus" or whatever qualifies as the minimal state of full belief. Replacing 1 by this assumption allows Fuhrmann to embrace assumption 2 and avoid my argument provided that the true theory T is accessible from the result of contracting K by contracting p. In practical and local deliberations, $\sim p$ is ruled out as a serious possibility, but when deliberating relative to ultimate goals, it is not ruled out.

I think of the belief-doubt model with assumption (1) as a consequence is the central contribution of the classical pragmatists. Messianic Realism is expendable as I think Dewey discreetly recognized. Dewey gave up on truth or avoidance of error as a feature of the goals of inquiry. I defend the secular (née myopic) realism adumbrated in Peirce and James that avoidance of error is a desideratum of the *proximate* aims of specific inquiries.

Embracing the double standard is in conflict with this original and compelling insight of the classical pragmatists. To save Messianic Realism by abandoning this insight is in my judgment to adopt the wrong set of priorities for epistemology. James may, indeed, have followed this route as Fuhrmann suggests. So much the worse for James.

KITCHER AND THE PRODUCTION OF RELIABLE INFORMATION

In retrospect, the title *Enterprise of Knowledge* that I used for my 1980 book articulates the concerns of at least two of Philip Kitcher's works better than it does my own. As he clearly states in "The Knowledge Business," Kitcher is interested in knowers and encyclopedias. Both are sources of information that are authoritative. They are, as such, of importance to inquirers (whether they are persons or scientific research teams or other research collectivities) who are consumers of such information. Kitcher is interested in how institutions that produce allegedly reliable information are and ought to be organized to succeed in such efforts to meet consumer demand.

This is an important and interesting topic and one to which I have given relatively little attention. I cannot be labeled a rugged individualist as Kitcher labels me for my failure to attend to this matter. I suspect that he sees me in this light because I insist that the point and function of scientific institutions and other institutional sources of information is to furnish information that *inquirers* – be they persons or institutional agents – can use in addressing the problems they face. And such sources are useful to an inquirer only if the inquirer takes for granted information about the reliability of the source and the value of the information provided. There is an important point about the concept of reliability that it is important to keep in mind in this connection.

The reliability of implementing a program for forming a belief in response to testimony from a source of information is reliability in just this sense. It is the chance of the report (testimony) made being true given that testimony is elicited. That is to say, it is the chance of a belief formed in response to the implementation of such a program being true given that such a program has been implemented. Call that the chance of a true belief formation (true BF) conditional on an elicitation E.

It is important to understand that implementing such a program for routine expansion must be done by the inquirer while the inquirer is in ignorance of what the content of the testimony yielded by the program (such as the information obtained by consulting an expert) happens to be. As long as the inquirer does not know the content of the testimony or any other relevant information, the inquirer can ground a judgment of credal or subjective probability is r that the belief that will be formed is true from the information that the chance of a true BF on a trial of kind E is r and the information that a trial of kind E has been implemented. Relative to that information, the inquirer can reason that running the program will import false belief with probability $1 - r$.

However, if the inquirer finds out the content of the testimony before forming the belief in response to the program, the inquirer then knows that the trial of kind E yields a belief that h (let us say) (call this event Eh). Let the chance of a true BF on a trial that is both a trial of kind E and of kind Eh be s different that r. Let r be high and s low. It is an objective fact that the program is reliable. That is to say, the chance of a true BF on a trial of kind E is high. The program implemented only after belief that h is formed is unreliable. Given the inquirer's information, the inquirer cannot use the reliable program. And he or she should not use the unreliable program. So the inquirer should use neither.

I think that Kitcher agrees with this point given what he says about reference classes. But the ramification of this agreement is that adopting a reliable program can be allowed only if the inquirer does not know the content of the belief formed on a trial of kind E prior to implementation and is committed before finding out the content to forming a belief with the content delivered no matter what the content delivered among those in contention happens to be.

Now it is just this precommitment aspect of routine expansion that implies that routine expansion is conflict-injecting. X may initially be convinced that h and add g incompatible with h to X's stock of beliefs and end up with an inconsistent stock. I do not see this as an objection to using routine expansion. We must put up with consulting experts and the senses in spite of this egregious defect in routine expansion. But because relying on authorities as Peirce reminded us and on the testimony of the senses (as he failed to observe) can land us in inconsistency and because it is doubtful that one can extricate oneself from such difficulties by exclusive reliance on other programs for routine behavior, we need to do more than merely decide which authorities to use as part of the public repository for information, which kinds of information to store in such repositories, and how to organize stored information in an efficient and accessible manner. I suppose that Kitcher agrees. We need to worry about the other repository of information – the information yielded by observation and experiment.

But the point I mean to belabor is that inquirers cannot rely exclusively on the public repositories of information and the testimony of the senses. Inquirers – whether they are persons or institutional agents – need to form decisions as to what to believe and not merely decisions as to which oracles to consult. My rugged individualism, as Kitcher puts it, is predicated on this observation (among others).

Kitcher appreciates the need to accommodate critics of the public system of knowledge. Big brother (the public system of knowledge) should encourage mavericks by a suitably constructed reward system for knowledge production. We should avoid the anarchy of a free market in epistemic matters.

All of this misses the main point. The public system of knowledge is a tool to facilitate the conduct of individual inquiries. It helps us to obtain information that cannot be obtained easily in other ways. But it is not the only tool. Observation and experiment is another such tool whose connection to the public system is complex but is not entirely to be assimilated into the public system.

Above all, none of this is any substitute for the judgment of individual inquirers regarding (1) assessing the value of information, (2) the importance of problems for inquiry, and (3) the determination of how to extricate oneself from the inconsistencies that the public system inevitably produces.

The public system of knowledge is an important tool, as both Kitcher and I agree. But for me, at any rate, it is a dangerous tool that calls for independent criticism not merely from those who are privileged by the system but from anyone who finds the inconsistencies produced by the system intolerable.

I think the issues raised by Kitcher are of first-rate importance. In my own work, I have not attended to some of them as much as I should have done. But I am confident that were I to do so, I would take up the cudgels against the big science that Kitcher finds so admirable.

INFALLIBILITY AND INCORRIGIBILITY: REPLY TO BENGT HANSSON

I take the inquirer X's state of full belief or knowledge to be X's standard for serious possibility that distinguishes between serious possibilities subject to genuine or serious doubt and impossibilities where doubts are mere paper doubts. This thesis is equivalent to endorsing the following:

Epistemological infallibilism of the present doctrine: According to rational X at the current time or in the current situation t, everything X fully believes at t is true. There is no possibility according to X at t that current beliefs are false in any sense that implies that according to X at t, X is risking error or making mistakes in embracing such beliefs.

I subscribe to the Peircean account of belief fixation that insists that there is no need to justify the current state of knowledge. Only changes in states of knowledge (i.e., states of full belief or standards of serious possibility) require justification. Moreover, such changes are justifiable through inquiry. Consequently, I endorse the following:

Corrigibilism of the present doctrine: According to rational *X* at *t*, it is seriously possible that *X*'s current state of full belief will be changed for good reason in the future by a sequence of expansions and contractions.

Corrigibilism of the present doctrine is equivalent to the doctrine of fallibilism of the future doctrine.

Fallibilism of the future doctrine: According to rational *X* at *t*, it is possible that *X*'s current state at time *t* will be legitimately transformed by a sequence of expansions and contractions so as to import a belief that *X* is currently certain is false.

Fallibilism of the future doctrine (i.e., corrigibilism of the present doctrine) is compatible with epistemological infallibilism. So are various contrary theses (which I do not endorse), such as the thesis that it is possible according to *X* that *X*'s current state of full belief will be changed for good reason in the future by a sequence of expansions without contractions, but contractions will never be legitimate. Or future contractions may be justified but expansions will never be. And there are other contrary theses. All of them are theses about the legitimate changeability of future states of full belief or standards for serious possibility.

So Hansson is right to claim that, on my view, corrigibilism of the present doctrine entails epistemological infallibilism, as do the various contraries.

I think Peirce agreed with me in all of this. When Peirce declared himself a fallibilist, I interpret him as a fallibilist concerning future doctrine – that is, as a corrigibilist about current doctrine. And I also understand him as maintaining that *X*'s current state of full belief is *X*'s current standard for serious possibility so that epistemological infallibilism is endorsed as well.

In *Enterprise of Knowledge*, I alleged that Peirce also seemed to endorse a third doctrine incompatible with the endorsement of the conjunction of epistemological infallibilism and corrigibility of the current doctrine:

Categorical fallibilism: Every rational agent should judge all logical possibilities to be serious possibilities in all deliberations and inquiries.

Categorical fallibilism in conjunction with epistemological infallibilism entails that the standard for serious possibility should be kept fixed as long as there is no change in the conceptual space. Corrigibilism of the current doctrine is rejected.

As I mention in my remarks on Misak's contribution, I based my claim that Peirce was a categorical fallibilist on passages that I read out of context. Misak was and is right to have objected to this interpretation of his views. I also agree with Misak that categorical fallibilism is incompatible with the most important elements of Peirce's view. Peirce did, however, endorse a conception of the ultimate goals and values of scientific inquiry – his "Messianic Realism" as I have called it – that either requires abandoning corrigibilism of the present doctrine in favor of categorical fallibilism or abandoning the thesis that X's state of full belief or knowledge is X's standard of serious possibility (i.e., the infallibility of the current doctrine). Peirce, so it seems, was inconsistent.

According to Messianic Realism, the ultimate goal of inquiry is to identify the true complete story of the world. The aim is messianic because for Peirce no such complete view can be uncovered in historical time. Messianic Realism is compatible with allowing that the solution to specific problems may be uncovered in historical time. But it entails that an inquirer focusing on a specific problem should seek to avoid importing false belief not only at the next change in doctrine but in any finite or infinite sequence of changes. And more crucially, if the inquirer is currently certain that h is true, X should avoid incurring any risk of following a sequence of changes in belief that replaces h with $\sim h$, never to return to h. From X's current point of view, such a possibility not only imports a belief that X is certain is false but fails to rectify the mistake in subsequent inquiry.

Fallibilism concerning future doctrine recognizes such sequences as serious possibilities. If X cares about avoiding error in the long run (as Messianic Realism requires) or in any number of changes after the first, X should never give up any current full belief and should not use any method of expansion that *might* mandate contraction at subsequent stages.

If X's current state of full belief is X's standard for serious possibility, X's current state ought never to be changed by contraction and, as further argument would suggest, by expansion either. The corrigibility of the current doctrine would be refuted. This is not quite the same as claiming that the current state of full belief ought to be the set of logical and conceptual truths, as categorical fallibilism requires. But that conclusion seems compelling once one recognizes that the alternative is to allow each inquirer his or her own incorrigible state of full belief without the hope of resolving differences between views through inquiry.

The unpleasant alternative is to give up the thesis that *X*'s state of belief is *X*'s standard for serious possibility. This leaves us in the dark as to what a state of full belief is for.

Bengt Hansson claims that on my view, corrigibilism of the current doctrine entails infallibility of the current doctrine. This is not true. If *X*'s current doctrine does not serve as *X*'s standard for serious possibility, the current doctrine can be changeable. I, to be sure, insist that epistemological infallibilism be maintained. So, I think, did Peirce. If *X*'s current doctrine were not *X*'s current standard for serious possibility, it is unclear to me what the point of having a current state of belief would be.

It should be emphasized here that for me, "fallibilism" is a highly ambiguous term. I have not begun to identify here all the different construals. (Some alternatives are canvassed in *Enterprise of Knowledge*.) In this discussion, I have contrasted categorical fallibilism with fallibilism of the current doctrine and fallibilism concerning future doctrine.

If I understand Hansson correctly, he takes the position that one should never be maximally certain concerning extralogical propositions. So what I called categorical fallibilism should be endorsed. Extralogical propositions vary with respect to degree of certainty (or probability). They also vary with respect to a factor that Hansson calls robustness. It looks rather like the kind of factor that Keynes called "weight of argument." Peirce had discussed a similar idea at great length as part of the "conceptualist" view of probability that he rejected and, indeed, rather fiercely opposed. (My own opposition is more qualified. I object to the insistence on numerically determinate probabilities required by classical Bayesian authors and to the intellectual imperialism and reductionism that probabilists advocate.)

When *X* is uncertain concerning which of rival alternatives constituting an ultimate partition is true, the rival alternatives and their Boolean combinations can be evaluated with respect to how probable they are on the evidence – that is, the current state of full belief or certainty. They can also be evaluated with respect to the value of the information they provide. Given an assessment of their probabilities and informational values, one is in general in a position to determine the degree of boldness needed to reject elements of the ultimate partition at a given level and their Boolean combinations. If the level of boldness is sufficiently low, there may be no point in further inquiry. The available evidence is sufficient to warrant converting an erstwhile conjecture that is initially judged possibly false into a full belief or certainty. By "certainty" here I mean *maximum* and *absolute* certainty. Absolute certainty is to be contrasted with probability 1 judgments that are *almost* certain. I contend that the inquirer may be justified in expanding his or her state of maximum certainties by adding new ones. I am here attending to *changes* in states of absolute certainty. I do not know whether this amounts to attending to dynamic aspects of belief change in a sense that Hansson would countenance. For me this change is a product of the inquirer's decision and not a response to an external "input." Such coming to full belief or certainty also takes place in response to external inputs such as the testimony of the senses or of witnesses and authorities. But the question of sufficiency of evidence to warrant ceasing further investigation arises in the context of inductive expansion. I agree with Hansson that such "robustness" is at least partially independent of degree of probability. It is a function of probability and informational value (see, for example, my *Gambling with Truth*).

I also agree with Hansson that degrees of corrigibility or degrees of entrenchment are independent of degrees of certainty. But degrees of incorrigibility or entrenchment offer us fine-grained discriminations between items that are maximally certain with respect to vulnerability to being given up. Degrees of certainty discriminate between propositions whose truth and falsity alike are serious possibilities. None of this has much to do with the contrast Hansson makes between degrees of certainty and degrees of robustness, both of which seem to be fine-grained discriminations among serious possibilities.

Hansson argues against my contention that talk of maximum certainty should be taken literally and not as a figure of speech by considering my example of a coin flying out to Alpha Centauri. This is not a serious possibility. Its negation is a maximum certainty according to my view and, so I think, to any one of good sense. Hansson maintains that claiming that the coin is flying out to Alpha Centauri at a speed approaching light is even less of a serious possibility because it is more specific and its negation is even more certain. But for the person, like me, who is certain that the coin is not flying out to Alpha Centauri, adding the information that it is not flying out at the speed of light is adding no information not already implied by the initial state of certainty. I agree that one can distinguish between propositions that are all equally maximally certain with respect to vulnerability to being given up. I agree also that such degrees of corrigibility are determined by assessments of losses of informational value incurred. Issues of greater or lesser specificity are relevant in such a setting. But none of this takes away from the fact that all full beliefs are equally and maximally certain even though they differ in the permanence of this certainty.

Hansson's main objection arises because he thinks that degrees of infallibility (certainty) and degrees of robustness (incorrigibility) can vary independently and that an extralogical belief can be sufficiently certain to count as if it were maximally certain and likewise sufficiently incorrigible to count as if it were maximally incorrigible. This runs counter to my view.

What is it for a proposition to be less than certain but sufficiently certain? Sufficient certainty must be for some purpose. If the purpose is not a practical one, it is, I suggest, sufficiently certain to have its status changed to that of a maximal certainty.

What is it for a proposition to be sufficiently incorrigible? Here too we must ascertain the purpose. I claim that we judge a proposition *already maximally certain* to be sufficiently incorrigible when it remains in the state of maximal certainties when a given proposition is required (for whatever) reason to be removed from the state of full beliefs. The example of Pb-206 illustrates this point nicely. It is maximally certain and sufficiently incorrigible. That is no problem, for degrees of incorrigibility are found among the maximal certainties. Because it is maximally certain it is, of course, also trivially the case that it is sufficiently certain.

WHEN INCONSISTENCY IS EPISTEMIC HELL: REPLY TO BUENO

Ottavio Bueno rightly understands that I do not urge avoiding inconsistent belief systems at all costs. I do not recommend avoiding programs for routine expansion via observation or the testimony of witnesses because inconsistency may be injected into the inquirer's state of full belief.

Nonetheless, I think that states of full belief that harbor contradiction are states of epistemic hell. Inquirers should make provision for retreating from them should they find themselves inadvertently in inconsistent belief states. An inconsistent state of full belief fails as a standard for serious possibility. In such a state all potential answers to the question under investigation are rejected. The challenge is to provide a rationale for "contraction" from an inconsistent state of full belief – not to learn how to live in an inconsistent state.

Bueno seems to disagree. But I am not sure how serious the disagreement is. I am speaking of states of full belief that are used as standards for serious possibility where all full beliefs are judged not only true but certainly true with no serious possibility of being false. Bueno contends that taking belief systems seriously does not require taking them as true. I agree that we can contemplate conjectures for test and development without taking them as true. But to fully believe that h is to judge that h is true and to do this for sure. It is not to conjecture that it is true.

Bueno thinks that "pursuing inconsistent belief systems" is a "useful device" and that my pragmatic approach would benefit from taking inconsistent belief systems and paraconsistent logic seriously. He claims that sometimes the only way to obtain information is by expansion into inconsistency.

Expansion of a state of full belief K by adding h is a case of adding the judgment that h is true to the set of judgments that are consequences of K. If K entails $\sim h$, the inquirer has deliberately expanded into inconsistency. Bohr's model of the hydrogen atom is incompatible with classical electromagnetism. According to Bueno, inquirers would not have obtained the explanations afforded by the new model and would have been deprived of valuable new information without such deliberate expansion.

I do not agree. An inquirer might reason about the Bohr's model hypothetically. That is to say, an inquirer need not fully believe that the Bohr model is true but merely *suppose* for the sake of the argument that it is true even while fully believing that classical electromagnetism is true. The inquirer need not expand his or her state of full belief into inconsistency at all.

It is also possible for an inquirer to contract his or her state of full belief so as to allow the inquirer to recognize Bohr's model to be a serious possibility. And if through suppositional and belief-contravening reflection, Bohr's model has come to be regarded as carrying sufficient informational value to provide incentive for such contraction, contraction of the initial state of full belief may be warranted.

I fail to see why it is necessary to introduce notions of quasi truth in order to address this matter. There is no need to contract an inconsistent belief system except in the case of inadvertent expansion into inconsistency.

The case of the inconsistencies in Frege's logicist program raise different issues. In the *Enterprise of Knowledge*, I disclaimed any attempt to account for changes in mathematical or logical knowledge.

UNCERTAINTY AND RISK: REPLY TO SAHLIN

Nils-Eric Sahlin has as an acute sensitivity to the moral and visceral motivations that have driven my efforts to construct an account of decision making alternative to the orthodox Bayesian view. I conjecture that this empathy derives from the efforts that he has made on his own as well as in conjunction with Peter Gärdenfors to elaborate on themes of risk, uncertainty, and ignorance similar to those that drive my own efforts and that I illustrated in my discussion of risk assessment in evaluations of the safety of nuclear power plants.

In the past, I have tended to emphasize my differences with others who have pioneered in promoting the growing rebellion against orthodox Bayesianism and probabilism. In recent years, however, I have come to appreciate more and more the affinities among my views and ideas and those of Gärdenfors and Sahlin, Ellsberg, Fine, Kyburg, Schick, Walley, Weichselberger, Sen, Jaffrey, and others who in one way or another have contributed to the understanding of indeterminate probability and utility. I have substantial differences with these authors as well as with my friends at Carnegie Mellon – Seidenfeld, Kadane, and Schervish. I can wax passionate about these differences – precisely because these disagreements are far more interesting to me than mere Bayes-bashing. This is true about Sahlin's work on what he calls "epistemic risk," including the development of the notion of evidentiary value of evidentiary mechanisms that Halldén pioneered along with Sahlin, Edman, and other members of the Lund school.

Sahlin has downplayed his differences with my views in his discussion so I shall mention only one respect in which his views and mine diverge. For me, given indeterminacy in probability and utility judgment, the primary constraint on rational choice is the principle of E-admissibility. To be rationally choosable, an option must be optimal in expected utility according to one permissible probability-utility pair. Sahlin's approach does not impose this constraint. Instead, Sahlin tacitly proposes another primary constraint – the constraint Sen and Walley call "maximality." No option may be choosen if there is at least one option to which it is inferior in expected utility according to every permissible probability-utility pair. All E-admissible options are maximal, but there are contexts where the converse fails. Typically, they involve situations where a choice is made between more than two options. And they illustrate situations where choice consistency situations including Luce and Raiffa's can fail.

Those who defend the requirement of E-admissibility and those who defend maximality, such as Ellsberg, Sen, Walley, and tacitly Sahlin, can invoke secondary criteria such as P-admissibility and S-admissibility. The cutting edge of current discussions of decision making under uncertainty, in my opinion, ought to be, and sometimes is, concerning the merits of requiring E-admissibility or maximality as primary. I have written on this issue elsewhere and shall not rehearse that discussion here. I merely wish to say that Sahlin has given fine expression to the alternative view.

The controversy concerning E-admissibility and maximality is of first-rate importance. But it could not be appreciated as a debatable topic without first recognizing the importance of indeterminacy in probability and utility judgment. Sahlin's own original contributions to this topic and his discussion of my views on risk assessment in the chapter to which I am responding are excellent appreciations of this importance.

PROBABILITY LOGIC AND LOGICAL PROBABILITY: REPLY TO KYBURG

I long ago abandoned an exclusively dispositional account of belief and of credal probabilitiy judgment. Dispositions cannot be taken to be doxastic unless they are understood to be fulfillments of doxastic commitments underwritten by principles of a logic of full belief and a probability logic. I plead innocent to Kyburg's allegation that for me, "the agent's beliefs, or set of dispositions to act come first, and is examined for conformity to certain standards before being given the honorific 'rational.' "

The norms and prescriptions of probability logic (or inductive logic, as I, following Carnap, used to call it) apply to an agent's *confirmational commitments*. A confirmational commitment is representable by a function from potential states of full belief to credal states. A credal state is representable by a set of conditional probability functions meeting formal requirements like those spelled out by De Finetti or Dubins or sketched by me in "On Indeterminate Probability" and *Enterprise of Knowledge*. By attributing a confirmational commitment to a rational agent, I am claiming that a rational agent X's is and ought to be committed at any given time to a method of judging what X's credal state ought to be relative to each potential state of full belief. So I think that X's state of credal probability judgment is uniquely determined by X's state of full belief and X's credal state. In this sense, X's state of credal probability judgment ought to be based on X's "evidence" – that is, X's state of full belief – in conformity with X's commitments as to what constitutes evidential support and, of course, in conformity with the requirements of the logic of full belief and probability logic.

Kyburg advocates a far stronger thesis. Probability logic ought to impose a system of constraints on rational agents so that everyone agreeing on the same "evidence" should as a matter of probability logic agree in their judgments of probability or in their partial beliefs. Probability logic requires that all rational agents ought to conform to the same standard confirmational commitment. I claim that confirmational commitments are subject to revision through critical inquiry. Kyburg insists that they are incorrigible. This has little to do with whether the cart or the horse comes first but with how demanding the horse of rationality is.

The requirements for a *Basic Bayesian Probability Logic* are listed below.

Confirmational consistency: If K is consistent, $C(K)$ should be representable by a nonempty set B of probability functions of the type $Q(x/y)$ where x is any proposition in the algebra of propositions under consideration and y is any other such proposition consistent with K.

Confirmational coherence: Every member $Q(x/y)$ of $B = C(K)$ should be a finitely additive conditional probability measure where $Q(x/y) = Q(x'/y')$ for every x, x' such that $K|{-x} \equiv x'$ and y, y' such that $K|{-y} \equiv y'$.

Confirmational convexity: If distinct Q and Q' are in $B = C(K)$, then for every e consistent with K, for every a between 0 and 1 inclusive, there is a Q^* such that $aQ(x/ep) + (1 - a)Q^*(x/e) = Q^*(x/e)$ is in B.

Confirmational conditionalization: Given potential state K and its expansion K_e^+ by adding e consistent with K, every Q_e^+ in $C(K_e^+)$ there is a Q in $C(K)$ such that $Q_e^+(x/y) = Q(x/y \wedge e)$ and for every Q in $C(K)$ there is a Q_e^+ such that $Q_e^+(x/y) = Q(x/y \wedge e)$.

Strict Bayesians would complain that confirmational convexity is too weak. It should be strengthened to the following:

Confirmational uniqueness: $C(K)$ should be a unit set. Credal probability judgment should be numerically determinate.[1]

Some authors who are plausibly counted as embracing a Bayesian probability logic (such as Teddy Seidenfeld) are critical of confirmational convexity but endorse some form of the other principles of basic Bayesian logic. Confirmational consistency, coherence, and conditionalization may be called principles of *core Bayesian logic*. Henry Kyburg is critical of confirmational convexity as well. By any stretch of an imagination grounded in the history of probability judgment, Kyburg cannot be considered to be a Bayesian. For he rejects confirmational conditionalization as well.

Given core Bayesian logic, a confirmational commitment is uniquely determined for all expansions of a minimal belief state *LK* by specifying an a priori credal state relative to *LK* each of which specifies a subset of the set of logically permissible conditional probability functions relative to *LK*. *The* logical confirmational commitment is representable by the set of all logically permissible conditional probability functions relative to *LK*. Other subsets represent extralogical confirmational commitments.

Thus, the logical probability or confirmational commitment is determined by probability logic, but once indeterminacy in probability judgment is allowed, there are extralogical confirmational commitments as well. Insisting on the adoption of the logical confirmational commitment is endorsing a form

[1] I have insisted on distinguishing between imprecise probability that is a question of measurement and indeterminate probability that is not (see "On Indeterminate Probabilities"). In discussing my view, Kyburg seems to gloss over this distinction.

of probability skepticism akin to the sort of Parmenidean view generated by insisting that full belief be restricted to logical truth.

This picture is predicated on endorsement of confirmational conditionalization. Kyburg rejects confirmational conditionalization. And he insists that although he allows a form of indeterminacy, the logical confirmational commitment is the uniquely rational one to adopt.

Yet both Kyburg and I agree that credal states need not be updated by temporal credal conditionalization. Kyburg thinks so because confirmational conditionalization is violated according to his account of probability logic. I think so because an inquirer may have good reason to modify his or her confirmational commitments.

Why Kyburg and I differ over confirmational conditionalization concerns differences in the way we think credal states are grounded in statistical information according to direct inference – especially differences concerning the selection of a reference class. The issue is discussed in *Enterprise of Knowledge*. I have learned more about this philosophically and methodologically important problem from disagreeing with Henry Kyburg than from anyone else. I shall always be grateful for his instruction.

SOME CHANCE DIVERSIONS: REPLY TO MELLOR

My old friend Hugh Mellor manages to articulate more interesting and important disagreements per page than I can ever keep up with. Even when he plays them down, each one is worth spilling more ink on than I can afford to do here. I will not even try. Here are comments on a few issues.

Mellor alleges that I think that a principle of direct inference tells us all we need to know about how evidence should effect our credences. He cites my contention that an objectivist inductive logic is a complete inductive logic in support of this claim. I continue to think that objectivist inductive or rather probability logic as I now prefer to say is a complete probability logic. But I never thought that a complete probability logic "tells us all we need to know about how evidence should affect our credences."

Mellor takes for granted here that a complete probability logic should yield a specification of a function from evidence to credences that represents a standard that all ideally rational agents should use in determining what their probability judgments ought to be given their evidence. Ever since I abandoned a youthful enthusiasm for Carnap's confirmation theory, I abandoned the idea that there is a single confirmational commitment (numerically determinate or indeterminate) to which all rational agents ought to subscribe.

My picture runs roughly like this. An inquiring agent X is in an epistemic state at t consisting of a state of full belief $\boldsymbol{K}$ and a confirmational commitment C. These taken together uniquely determine X's state of credal probability judgment at t. X's state of full belief is subject to modification over time *and so is X's confirmational commitment*. Epistemology as I think it ought to be done should concern itself with indicating the kinds of considerations relevant to modifying states of full belief and confirmational commitments.

I once alleged that Mellor's view of chance introduced some gratuitous metaphysics. Later on I qualified my view by saying that either his conception was gratuitous or it committed hostages to a vision of objective chance that is incompatible with determinism. On the view I favor, it is the coin that has the stochastic property – the chance. The stochastic property, like dispositional properties, is relativized to trial types and outcome types. X has a disposition to dissolve on immersion in water. The coin has a 50 percent chance of landing heads on a toss. In direct inference, the stochastic predicate is alleged to be true of some object. It is also alleged that a trial of some type is performed on the object. Some judgment of credal probability is grounded in this information. My complaint was that Mellor broke up this sequence of steps into additional steps. In lieu of the chance predicate true of the coin, the coin had a sure-fire disposition to yield a 50 percent heads/50 percent tails chance display on a toss. Direct inference was then to be made from the information about the chance display to the credences.

Now if the chance display is a feature of the outcome of tossing (the landing heads up, tails up, or on its edge), there is a problem concerning the relativity of chance to the kind of trial. This relativity so I would argue is crucial to the way one can accommodate objective chance with an underlying determinism pace James Bernoulli. If there is no relativity, there is no problem of the reference class. If, on the other hand, such relativity is introduced, then it seems to me that locating the chance as a feature of the outcome of tossing or as a feature of the tossing is gratuitous and diversionary. It is gratuitous because it is unnecessary and diversionary because once one has the notion it has a life of its own and can become a separate object of study.

At one place, Mellor seems to be responding to my allegation of gratuitousness. There are no references to chance displays and dispositions to such displays here. The allegation of gratuitousness surfaces in his distinction between a's having a chance and a toss of a having a chance. The coin a has a chance p of landing heads in the sense that it supports the conditional that if a were tossed endlessly with each toss having a chance p of landing heads, the limiting frequency of heads would be p.

I would prefer in this case to suggest that it is the coin that has a chance at t of landing heads up on a toss. I would then say that if a were tossed endlessly and it retained the same chance p of heads on a toss during the entire process, the limiting frequency of heads would be p. Strictly speaking, this is not right. It would be better yet to say that throughout the entire period the coin has the following family of stochastic properties – to wit, on a trial consisting of n tosses, the *joint* chance distribution over the space of possible outcomes has itself the following properties: (1) the marginal chance of heads on each toss is p, and (2) tosses are independent. On these suppositions, the limiting frequency of heads would be p.

Mellor, like so many philosophers, seeks to build up a series of repetitions of a kind of trial using a single conception of tossing a coin. To get the notion of the marginal distributions being "identical," the chance is predicated of tosses. This still leaves the question of the features of the "joint" distribution such as probabilistic independence to be considered. Attributing chances to tosses does not help here. We need to attribute them to outcomes of n-tuples of tosses, and so on.

I prefer to think of chances as true or false of coins relative to a specification of kinds of trials and sample spaces of outcomes – as is the practice in statistical analysis. I do not pretend that the metaphysics I deploy in characterizing this practice is perfect. There are, perhaps, respects in which my account may have the gratuitous excesses squeezed out of it. My challenge to Hugh remains as it was: how to squeeze the excesses out of his account in a fashion that respects statistical practice. Perhaps a common effort focused on the topic of direct inference might lead both of us to the kind of consensus that we both could enthusiastically endorse.

I write all this because I am, in spite of my suspicions about the excesses of contemporary metaphysics, ready to take any theoretical issue seriously insofar as it contributes to the understanding of the conduct of inquiry and its products. I believe that Hugh shares the same attitude. The devil lurks, as Hugh points out, in the details.

CORRIGIBILISM AND NOT SEPARATISM DIVIDES US: REPLY TO SPOHN

I concede to Wolfgang Spohn that some of my past efforts to characterize his views are not consonant with his intention. I believe that a more recent attempt I have made (in a talk to be given in Konstanz) comes closer. But if we are to credit Spohn's remarks in his intriguing chapter, I still haven't quite got the point.

Spohn declares himself a separatist concerning the relations between belief and probability. I guess I have been reluctant to take his previous hints (not so persistently presented as here) as seriously as I should have. To take them seriously is to interpret Spohn as engaging in a grand refusal to engage in an effort to find systematic connections between belief and probability. As he writes here:

> Separatism is the view that acknowledges a useful theory of subjective probability and a useful theory of belief, and keeps them as coexisting, but separate enterprises not in need of unification in an integral picture.
>
> Whether I end up as a genuine separatist is not so clear. I would like to see myself as at least formally returning to reductionism . . . ; this would be theoretically most satisfying. Soberly, though, I prefer to present myself as a separatist.

I still find it difficult to take Spohn at his word.

In the first place, he himself does make gestures at integration in several places as in the appendix of his excellent chapter. As he himself explicitly acknowledges, there is a near algorithmic way of linking his method for updating ranking functions with Jeffrey updating.

Spohn takes himself to be interested in a univocal notion of belief. He proposes to explicate that notion in terms of his ranking function. If $\kappa(\sim h) > 0$, h is believed. This readily translates into the Shackle terminology as claiming that h is believed if and only if h is believed to a positive degree – that is, $\sim h$ is surprising to a positive degree.

Spohn objects to my imposing my interpretation of Shackle measures on his account. According to that account, when the inquirer X is in belief state K relative to which his or her degree of belief that h is positive, if X is maximally bold, he or she is justified in expanding K by adding h. Once X has expanded by adding h to K, his or her degree of belief that h is not merely positive. X changes X's belief state by becoming certain that h.

There is another interpretation with which he has flirted that he elaborates in the appendix to his chapter. Degrees of disbelief represent the "ranks" in "ranked probabilities" where the ranked probabilities can be equated with a ranked sequence of unconditional probabilities that can, in turn, be equated with a Popper measure.

Although he has discussed this idea in various places and continues to do so, Spohn explicitly disavows it because it leads to the view that only propositions carrying probability 1 may be considered plainly believed.

Why then does he assign this nonarchemidean view any more significance than my proposal for building bridges between κ and other concepts?

The reason seems clear, and it has little to do with being interactionist or separatist. Spohn is as resolute as R. C. Jeffrey and the radical probabilists in refusing to recognize the question of changing states of full belief or states of certainty as an epistemological problem. I do not mean to suggest that Spohn is a radical probabilist. As he rightly notes, probabilists are eliminativists when it comes to plain belief. Neither Spohn nor I are. (Plain belief was called "mere belief" in *Gambling with Truth* and "positive belief" elsewhere.)

On my view, X plainly believes that h if and only if X currently is prepared to reject $\sim h$ via inductive expansion should X be maximally bold. Note that in X's current state, X has not rejected $\sim h$ and may not be prepared to reject $\sim h$. X may not be bold enough. But if X were maximally bold, X would or at least should be prepared to change X's state of full belief by adding h to the stock of full beliefs.

Spohn must find this intereactionist proposal objectionable not because it is interactionist but because it makes the importance and usefulness of plain belief depend on its role in expansion of states of full belief.

Spohn maintains that X should never change X's state of full belief. So either Spohn requires X to be minimally bold and insists that X add h if and only if $\sim h$ is maximally surprising or $\kappa(\sim h) = \infty$, or he refuses to countenance my way of allowing plain belief (and degrees of disbelief) to interact with probability. Spohn's separatism amounts to his adoption of the second option.

The idea that degrees of disbelief identify the ranks of ranked probability functions does not presuppose for its applicability that states of full belief may legitimately change. Even though this application requires X to assign probability 1 to plain beliefs, it does not require X to fully believe them. So they are revisable according to Spohn's principles.

Spohn in the end abandons the ranked probability construal of disbelief. His reason is that it is not plausible that all plain beliefs are certain.

Spohn's distaste for changes in full belief is expressed also in his view of "updating" κ functions. As Spohn has observed, his techniques parallel Jeffrey updating. Jeffrey updating avoids changing states of full belief – or so it is alleged. I myself am on record as regarding Jeffrey updating as poorly conceived as an account of how to modify probability judgments through the making of observations. Inquirer X's credal probability judgment is determined by X's state of full belief according to X's confirmational commitment. The credal state may change in virtue of changes in the state of full belief or the confirmational commitment. If the confirmational commitment remains constant, the credal state changes only if the state of full belief or evidence changes.

Jeffrey supposed that it could change in response to sensory input. This is so when the confirmational commitment is held constant and the state of full belief is expanded in response to sensory stimulation according to some program that the inquirer is committed to judge reliable. But Jeffrey denied that the response to sensory stimulation involves expansion of the state of full belief. Moreover, the response could not be judged reliable since such a judgment would invoke full belief. I confess I do not understand what Jeffrey was talking about and find the difficulties I raised back in 1967 as compelling now as I did then.

I do not think the transformation of Jeffrey update of credal probabilities into Spohn update of ranking functions in response to observation improves the idea. Spohn thinks that moving from the claim that the subject directly perceives that *A* to the claim that "experience affects the subject's beliefs somehow" is a mark of philosophical progress. I do not. How an inquirer could use Spohn's recipe for updating for exercising critical control over the updating process remains obscure. Presumably, such critical control is not to be had. To my way of thinking, the price for refusing to acknowledge that full beliefs are subject to change is much too high.

I admire and respect the seriousness of Spohn's project and the skill and ingenuity he has exercised in developing it. I appreciate Spohn's departure from the narrow confines of radical probabilism. In the end, however, it is no more acceptable.

ABDUCTION, INDUCTION, ENTRENCHMENT, AND SURPRISE: REPLY TO PAGNUCCO

I am most grateful for Maurice Pagnucco's generous, fair, and accurate description of the understanding of abduction and induction according to Peirce, myself, and scholars working on artificial intelligence.

As Fann noted a long time ago, Peirce did indeed change his views on this topic – so much so that he felt obliged to alter his terminology. He did not begin using the term "abduction" until the beginning of the twentieth century around the time that he explicitly disavowed his earlier conception of the distinction between what he then called "hypothesis" and induction.

Pagnucco is right to point out that Peirce's earlier trichotomy among deduction, induction, and hypothesis covered many functions, some of which were better collected under induction. Peirce also tended to write in a manner that surely suggested that the form of ampliative inference he called "hypothesis" is what writers since G. Harman have called "inference to the best explanation."

Pagnucco also took note of Peirce's emphasis on a formal contrast among the three kinds of inference. This contrast was important and, indeed, central to the thought of the young Peirce. At that stage in his career (and well before Frege), Peirce was emphatic in his opposition to psychologism in logic. According to his "unpsychologistic" logic, logic was a formal discipline. But in contrast to Frege, this formal discipline covered ampliative reasoning, including both hypothesis and induction.

Peirce was not so foolish as to suggest that one could provide accounts of the hypothetical and inductive validity of inferences in formal terms. He explicitly disavowed such intent. But he attached great importance to offering formal characterizations of the kinds of arguments that belong to hypothetical and inductive reasoning. And this feature of Peirce's thought has stimulated the efforts of workers in artificial intelligence interested as they are in the design of intelligent reasoning devices.

Peirce himself, I have conjectured, changed his own understanding due to issues that remained unsettled for him after he had made his most important contribution to inductive reasoning: to wit, a pioneering account of confidence interval estimation in the spirit of Neyman, Pearson, and Wald. In later work, Peirce called his proposal an account of quantitative induction. He had seen that this account could accommodate much of what fell under nonstatistical as well as statistical induction. But there are types of inference that he had been inclined initially to classify as hypothetical reasoning that he then appreciated could be regarded as induction. This led him to reinvent his contrast.

Abduction (alias hypothesis) became the proposing of conjectures for test and development (as my good friend Sidney Morgenbesser used to say). But unlike K. R. Popper, Peirce thought that one tested hypotheses in order to attempt to settle unsettled questions. Such testing would lead to the rejection of one alternative rather than another. Such rejection, according to the statistically minded Peirce, is almost always nondeductive or ampliative or inductive in character. To all intents and purposes, there could be no testing without induction or ampliative reasoning. Popper, by way of contrast, preached what seems to a Peircean to be the quixotic sermon that one should engage in abduction without induction. Other anti-inductivists have bowdlerized Peirce in yet another fashion.

Quine and Ullian suggest that induction is but a special limiting case of abduction. But deciding between conjectures on the basis of experiment and other information is not a limiting case of conjecturing.

Pagnucco is far more sensitive to the distinction between abduction and induction in the way the later Peirce conceived of it. And he has even more interestingly used this sensitivity to address some of the deficiencies in the

AGM approach to belief change. In particular, he has sought to provide an account of expansion that is more than adding information to a theory by stipulation and forming the deductive closure. Whatever the limitations of AGM contraction may be, an account is given of how to evaluate and choose between rival contractions removing some specific proposition from a theory or belief set. The choice is related to an evaluation of propositions with respect to entrenchment. In his dissertation of 1996, Pagnucco saw that one might try to extend the evaluation of sentences already in K with respect to entrenchment to an evaluation of sentences recognized as potential answers to questions at the abductive phase so that they are consistent candidates for expansion rather than contraction. Pagnucco's idea parallels a line of thinking already present in Gärdenfors and Makinson in 1993 and in Spohn in 1988. He adds to it the idea that the extended entrenchment ranking can be used as a satisficing measure for determining what to add inductively to the initial belief set.

This approach is based on what, in my judgment, is a half-truth. I have argued since *Gambling with Truth* that inductive expansion should be based on boldness-dependent and deductively cogent expansion rules. More specifically, I favor rules derived from the injunction to maximize expected epistemic utility where the epistemic utility function is a weighted average of a utility function $T(x, y)$ where x is a proposition added to K and y is t or f for truth value and $T(x, t) = 1$ and $T(x, f) = 0$ and a utility function $C(x, y)$ where $C(x, t) = C(x, f) = Cont(x) = 1 - M(x)$ and boldness is determined by the weight α greater than or equal to 0.5 and less than or equal to 1. But the approach I am now describing can be generalized to allow for other criteria. And in that book I showed that Shackle's measure of potential surprise and its dual measure for degree of belief can be derived from the resulting criterion for determining which propositions are added to the initial K. Indeed, any proposition may be added if maximum boldness is exercised ($\alpha = 0.5$) if and only if it carries some positive degree of belief. It then became widely recognized that entrenchment measures have formal properties similar to surprise measures. Spohn, Gärdenfors, Makinson, and Pagnucco make the additional substantial assumption that entrenchment is equivalent to surprise.

Pagnucco recognizes that doing so is a substantial assumption. It still seems unacceptable to me. Entrenchment as I conceive of it is a function of the informational value-determining probability M alone. Surprise is a function of the M-function and the degree of belief function. My contention is based on a vision of the aims of inquiry in expansion and in contraction. Pagnucco disagrees. Our differences over the aims of inductive expansion and how abduction contributes to promoting these aims have ramifications also for how we address iterated expansion and iterated contraction. Whatever the merits of either side of this dispute, the issues involved are of tremendous philosophical significance and apparently have relevance to topics in artificial intelligence.

ANSWERING BACK: REPLY TO OLSSON

Acknowledging one's ignorance is *always* a relevant response to a question of fact. If asked (in the summer of 2004) who will win the United States presidential election, saying, "I cannot answer the question" is answering the question.

There is no paradox here but only ambiguity. Questions requesting the addition of information to the inquirer's state of full belief K recognize as relevant linguistic responses any expression of a proposition to be added to K that is a Boolean combination of propositions that are strongest propositions whose addition to K gratify the demand for new information articulated by the question. Relevant linguistic responses express, in this sense, potential answers to the question being raised. Responses such as "I cannot answer the question" and "Bush or Kerry but not Nader will win the election" express relevant responses to the question and are also relevant answers.[2]

Linguistic responses expressing the strongest consistent propositions whose addition to K gratify the demand for new information are not only relevant linguistic responses, they are responses that, if added to K, render further inquiry into the question under study pointless – unless the demands for new information articulated by the question are reconsidered. "Bush will win," "Kerry will win," and "Nader will win" are answers in this sense.

Olsson does not dispute the assumption about cognitive options he formulates as (1) in his chapter on potential answers. He suggests, however, that (2) and (3), which relate cognitive options in inductive expansion with potential answers, fly in the face of common sense. This is so if we take potential answers to be answers that, if adopted, render further inquiry into the question pointless. But if we understand potential answers to be relevant responses to the question, cognitive values and, I dare say, common sense ought to find (2) and (3) congenial.

The appeal to intuition and high probability rules for inductive acceptance ought also to be resisted. I dare say that inductive expansion criteria can be formulated by stating that K ought to be expanded by adding to K all and only

[2] "Dean will not win" expresses the same relevant response as "I don't know" given that that K entails that Dean will not win.

propositions whose evidential support by K is above some given threshold. That there is a sense of evidential support according to which this applies ought to be noncontroversial.

That evidential support so construed satisfies the requirements of the calculus of probabilities is an additional and much more debatable assumption. Appeals to intuition to support it are just so much dogmatism. Indeed, the example of the lottery ought to prompt some doubt about the assumption, since it leads to the so called "lottery paradox."[3] Olsson points out correctly that high probability rules lead to inconsistency in the sense that the set of sentences added to K cannot all be true. This leads, as he notes to a trilemma. One of the alternatives is giving up such rules. Another is giving up the concern to avoid error. The third is to regard expansion of K to be epiphenomenal. One accepts highly probable propositions without adding them as evidence. Olsson correctly chastises Bovens and Hawthorne for following the third course where acceptance is not provided with useful application.

There are some elements of Olsson's account of my positive view that require modification. My account of inductive expansion in *Gambling with Truth* was modified in an important respect in "Information and Inference," written in response to the epistemic utility argument developed by Hintikka and Pietarinen in the early 1960s. Olsson's account of my account and how it bears on the lottery emphasizes an approach that belongs to my earlier approach.

In *Gambling with Truth*, I required that elements of the ultimate partition U_K relative to K carry equal informational value. This meant that in the case of the lottery, a "Hamlet question" where the hypothesis "ticket i will win" and its negation are elements of the ultimate partition, each element would carry equal informational value of $\frac{1}{2}$ and would be rejected if its probability were less than $q/2$ where q is the index of boldness. In order for "ticket i will win" carrying probability 10^{-6} to avoid rejection, q would have to be less than or equal to $2(10^{-6})$.

Kyburg and others complained that this approach made inductive expansion relative to the question raised and the ultimate partition. As long as inquirer X asks only one question with its associated ultimate partition, this should create no problem. But what is to prevent an inquirer from raising two or more questions with overlapping ultimate partitions or two or more

[3] This doubt may be reinforced by taking note of the fact that "satisficing measures of inductive support" that do not lead to the lottery paradox and can be given good motivation are available. See my "Maximizing and Satisficing Evidential Support," in *Reading Natural Philosophy,* D. Malament (ed.) (Peru, Ill.: Open Court, 2002), pp. 315–34.

investigators from coming to distinct conclusions because they ask two questions? X might ask, "Which ticket in the lottery will win?" and legitimately suspend judgment, whereas Y asks the "Hamlet question" about ticket 1 and comes to full belief that ticket 1 will lose. X and Y are in disagreement here. How can the dispute be resolved?

There are several issues that need addressing here. First, there is the "Hamlet question." According to the model of inductive expansion elaborated in *Gambling with Truth*, if an ultimate partition has two elements, an element x of the ultimate partition is accepted if and only if its probability $Q(x)$ is greater than $1–0.5q$. Here q is the index of boldness and takes values from 0 to 1. This is a high probability rule.

Now Olsson claims that I concede that in the lottery problem, it is possible to predict that ticket 1 will lose if the question is the Hamlet question: Will the ticket win or not? Olsson bases this on my discussion in *Gambling with Truth* where I go further than this and insist that relative to such questions, consistency of the inductive expansion rule *requires* that high probability (probability greater than 0.5) is necessary for acceptance. But this is precisely one of the respects in which "Information and Inference" *departs* from *Gambling with Truth*. Since writing that essay thirty-seven or thirty-eight years ago, I have disavowed the necessity of high probability even for Hamlet questions.

In *Gambling with Truth* I required all elements of the ultimate partition to carry equal informational value. Here the informational value of a potential answer h relative to K and the ultimate partition U_K is represented by $Cont_K(h) = 1 - M_K(h)$ and M_K is a probability measure defined for propositions equivalent given K to Boolean combinations of elements of U_K. Hence, if U_K contained n elements and h is the disjunction of m of them, $M(h) = m/n$, $Cont_K(h) = (n - m)/n$ and an element x of U_K is rejected at level of boldness q if and only if $Q(x) < qM(x) = qm/n$.

"Information and Inference" abandoned the assumption that all elements of the ultimate partition must carry equal M_K value. In the case of a million-ticket lottery, it is plausible that all hypotheses of the form "ticket i will be drawn" will be judged to carry equal informational value by most inquirers. This is not a claim about rationality but simply an observation to the effect that the interests of inquirers in relieving doubt would typically be of this kind. But in addressing the "Hamlet question," it is not, in general, plausible to think that ticket 1 will not win is evaluated as carrying the same informational value as the claim that ticket 1 will win. Although rationality does not prohibit evaluation of the two alternatives as equal, most individuals would judge the first alternative mentioned to be far less specific than the second and, hence, as carrying far less informational value. Indeed, "ticket 1 will win" would relieve doubt to the same degree for most people in the context of the Hamlet question as it does when the ultimate partition consists of all million alternatives of the form "ticket *i* will win."

When this is the case, in the Hamlet question about ticket 1, the high probability attributed to the prospect that ticket 1 will lose is not sufficient for acceptance. And if the proposition that ticket 1 will win were just slightly greater than 10^{-6}, it could be accepted if the agent were sufficiently bold. High probability is not necessary for acceptance *even in the case of a Hamlet question*.

So the acceptability of an inductive expansion is relative not only to the question asked (or the ultimate partition) but to the function M_K as well.

This additional relativity may seem to make the problem of what to do when multiple questions are raised simultaneously all the more urgent. In point of fact, it makes its solution much easier, as the discussion in "Abduction and Demands for Information" (chapter 7 of my *Decisions and Revisions*) indicates.

First, consider a case where there are two ultimate partitions. If we were seeking a consensus between two inquirers pursuing different questions, we should move to the coarsest common refinement. If *X* pursues the question as to which of a million tickets will win and *Y* asks whether ticket 1 will win, the ultimate partition that is the coarsest common refinement of the two is *X*'s ultimate partition.

Now if *X* and *Y* are already in agreement as to the informational value of "ticket 1 will win," it would seem that they should be willing to use *X*'s M_K function over the coarsest common refinement. The upshot is that they should end up agreeing to suspend judgment as to which ticket will win.

But suppose for the sake of the argument that *Y*'s assessment of informational value for the Hamlet question rates "ticket 1 will win" and "ticket 1 will not win" as carrying equal informational value of $\frac{1}{2}$, as *Gambling with Truth* recommends. In reaching consensus with *X*, the coarsest common refinement would be used. *X* would use the assessment according to which every hypothesis of the form "ticket *i* will win" carries M_K- value of 10^{-6}. But according to *Y*, "ticket 1 will win" carries M_K- value 0.5. The other hypotheses of the form "ticket *i* will win" carry values consistent with this in a manner that it is up to *Y* to specify. Given the specification, in consensus, *X* and *Y* recognize as "permissible" all M_K functions belonging to the convex hull of *X*'s and *Y*'s function.

Without going into details, it turns out that expanding by adding "ticket 1 will lose" and complete suspense are both E-admissible. According to my rule for cases where more than one option is E-admissible, the weakest

E-admissible option is to be recommended if it exists as it does. So we end up suspending judgment once more.

The same kind of result ensues if the consensus between the million Hamlet questions is sought.

I have spoken here of a consensus between several agents raising different questions. But I assume that a similar analysis applies to a single agent who is engaged simultaneously in all of these questions.

So even if a single agent asks Hamlet questions with assessments of informational value along the lines I favored in *Gambling with Truth*, high probability will be insufficient for acceptance as well as unnecessary.

Of course, an agent who asks just one Hamlet question and uses *Gambling with Truth* methods would consider high probability necessary for acceptance in that case. However, I wish to emphasize that in order to rationalize the high probability rule in that case, the informational value of adding "ticket *i* will not win" needs to carry informational value equal to adding "ticket *i* will win." I submit that few investigators would evaluate the value of information in this way (counter to the proposal in *Gambling with Truth*).

This brings me to the final issue raised by Olsson in his provocative chapter about the lottery.

In *Enterprise of Knowledge*, I embraced the idea that to claim that *X* knows that *h* is to claim that *X* fully believes or is certain that *h* and that *h* is true. I contend that this is so no matter who claims that *X* knows that *h*. If the person who attributes the propositional knowledge to *X* is *X*, *X* does so while fully believing that *h* and, hence, judging that *h* is true. So *X* is then committed to judging that *X* truly believes that *h*. Since *X* stands in no need of justifying *X*'s full belief, there is no basis for depriving *X*'s full belief that *h* of the honorific epithet of knowledge. If *Y* attributes belief that *h* to *X* and *Y* agrees with *X* that *h* is true, for *Y* to deprive *X*'s belief of the status of knowledge would be to suggest that there is something defective about *X*'s belief that suggests that *X* should cease believing that *h*, pending remedying the defect. Advocates of justified true belief or other pedigree epistemologies insist that *Y* should encourage *X* to remove belief that *h* from *X*'s standard for serious possibility if the pedigree of the belief is somehow questionable. I disagree and, indeed, take this view to be both petty and mean-spirited.

If I am right, the condition that *X* knows that *h* if and only if *X* truly believes that *h* ought to hold not only from *X*'s point of view, as Olsson rightly acknowledges, but from *Y*'s point of view as well.

Of course, I make this claim concerning the attributor's and *X*'s doxastic commitments. So I am focusing on what *X* believes insofar as *X* is fulfilling *X*'s doxastic commitments. So to claim that *X* knows that *h* may be glossed as claiming that any reasonable person in X's position (i.e., anyone fulfilling X's doxastic commitments) would know that h. I am not sure how close this is to Wittgenstein's intentions in the passage cited by Olsson and I do not have enough interest in Wittgensteinian hermeneutics to care. In any case, Olsson does not think my account is social enough. Not only must anyone in X's position share X's doxastic commitments, but, according to Olsson, there must be "grounds that are socially recognized as such" for X's doxastic commitments – in particular, X's belief that h.

Now I do concede that there is a social conception of what it is to be a "knower" – that is, an expert or an authority on some topic. A knower Y on some topic is someone whose opinions on the topic are reliable so that X may reasonably expand X's state of full belief by adding Y's testimony to X's full beliefs via routine expansion. It is in this sense that not only experts but scientific institutions and encyclopedias may be taken to be repositories of knowledge. It is reasonable to demand that such institutions and experts have proper credentials. And as Philip Kitcher rightly emphasizes, such credentialing has irreducibly social, political, economic, and moral dimensions to it.

Now Olsson claims that "when I say that I know, I am not just expressing my own certitude, I am also committing myself to the existence of grounds that are socially recognizable as such. I am giving others license to take the same view as I have."

When I say that I know, I do not for the most part give myself airs that I am authorizing others to take the same view as I have. In general, I do not think I have any right to license them to agree or disagree with me. There are, however, contexts where I am called on as a teacher or as a witness or authority to give expert testimony. If my credentials are that I report the deliverances of tea leaves, I would surely be suspect as an authority whose word may be taken as gospel truth.

When I say that I know that h, I am expressing my doxastic commitment. I am reporting my full belief that h. If you disagree, you may reasonably ask me to justify to you why you should come to agree with me. To do so, I need not show you how I came to have the conviction that h is true. I am under no more obligation to justify my conviction to you than I am to myself unless I am claiming expert status and that you should take me at my word.

Olsson thinks my view is mistaken and offers as evidence the observation that "while one may claim to be certain that one's ticket will lose, in the sense of excluding it as a serious possibility, it seems awkward to claim to know that one's ticket will lose."

I agree that it is awkward because it is awkward to claim or come to claim via inductive expansion that one's ticket in a fair lottery will lose. As

I have explained, even when asking the Hamlet question in such cases, one should not come to such conviction. I acknowledge that there is a system of epistemic values that licenses that result. But I contend that such values would be awkward from most points of view. In a fair lottery, the claim that one's ticket will win is held in suspense. Of course, knowledge is then out of the question. Olsson's counterinstance to my true belief account of knowing that (as opposed to being a knower) is no counterinstance at all.

INFORMATIONAL VALUE SHOULD BE RELEVANT AND DAMPED! REPLY TO ROTT

I very much appreciate Hans Rott's discussion of my decision theoretic approach to inductive expansion and to the question of how to contract. Rott has pioneered his own decision theoretic approach to belief revision and has developed along with Maurice Pagnucco the approach to contraction I now favor with a rationale alternative to my own. Rott's chapter is a response to ideas I presented in *Mild Contraction* and to my comments on his collaboration with Pagnucco. His remarks promote progress by enabling us to identify mistakes and remove misunderstandings. Exchanging ideas with him is unfailingly rewarding.

In note 23, Rott points to one of the reasons he continues to hold back from a firm endorsement of "severe withdrawal" alias "mild contraction." He worries that Sven Ove Hansson may be right in claiming that severe withdrawal is too "expulsive." Consider two sentences h and h' in inquirer X's corpus. Hansson alleges that according to severe withdrawal, X loses h' when removing h or loses h when removing h' whether or not h and h' are relevant to one another.

Hansson's observation is correct for Rott's version of severe withdrawal but not for mine. According to Rott's account, if both sentences h and h' are in X's corpus in language L, a maxichoice contraction removing h from the corpus K will be the intersection of K with a maximally consistent set in L containing $\sim h$. A saturatable contraction removing h will be the intersection of such a maxichoice contraction removing h with a subset of the set of maximally consistent sets in L containing h but inconsistent with K.

Severe withdrawal according to Rott and Rott and Pagnucco is defined in terms of maxichoice contractions so conceived. In that case, either removing h from K removes h' or removing h' from K removes h.

According to my proposal, potential contractions are restricted to belief states that represent relevant answers to the inquirer's question as represented by elements of a *basic* partition U_{LK} representing expansions of a minimal belief state LK that are incompatible with K. The elements of the basic partition need not be and in general would not be maximally consistent theories in L. They most surely would not be if L contains sentences h and h' that "have nothing to do with one another." If expanding LK by adding h is a potential answer, then some cells in the basic partition entail h and the remainder entail $\sim h$. But no cell in that basic partition entails h' and no cell entails $\sim h$. In an inquiry where h' is a potential answer, the situation is reversed. Consequently, a maxichoice contraction removing h from K is not the intersection with K of a maximally consistent corpus or theory in L but a maximally consistent potential or relevant answer in the roster generated by the basic partition U_{LK} incompatible with K or, alternatively, by a cell in the dual ultimate partition U^*_K. Saturatable contractions and mild contractions are understood in a similar vein.

In inquiry, the relevant or potential answers do not include all propositions expressible in the language used to represent belief states unless that language is restricted deliberately to the means for expressing such relevant answers. Basic partitions relative to LK that, together with K, determine dual ultimate partitions for contraction and ultimate partitions for expansion are designed to mark off the relevant potential answers to the question under investigation.[4]

Thus, Hansson's objection to mild contraction (or severe withdrawal) does not apply to contractions that arise in the context of an inquiry where the basic partition is given. It is true that given any pair of sentences h and h' entailed by K such that some element of the dual ultimate partition U^*_K entails h, at least one entails h', at least one entails $\sim h$, at least one entails $\sim h'$, and either removing h requires removing h' or vice versa. But it is not true in that case that h and h' have "nothing to do with one another."

Rott and I agree that in spite of the formal similarities in the recommendation of severe withdrawal and mild contraction, there are philosophically interesting differences in the rationales.

[4] Rott seems to have underappreciated this point. He discusses both deliberate expansion and contraction without benefit of the basic partition. When characterizing potential contractions, however, he seems to treat ultimate partitions as if they were basic relative to the corpus of logical truth. However, potential or relevant contractions cannot be characterized using the current corpus K and the ultimate partition U_K. One must either specify the dual ultimate partition independently or invoke the minimal corpus and the basic partition U_{LK}. Once these are in place, one can derive both U_K and U^*_k. Also I do not think that Rott's modification of my requirement that a dual ultimate partition contain at least one element entailing the negation of the item h in K to be removed is useful (note 17). Rott suggests that in this case, contraction removing h should be the identity transformation K. I prefer to say that h is not eligible for contraction in such a case. Notice, by the way, that if one does follow Rott's proposal, Hansson's objection against severe withdrawal would have to be modified.

I contend that the set of contractions available for choice that remove h from K should be evaluated with respect to loss of version 2 damped informational value. This evaluation provides a comparison that weakly orders *all* potential contractions removing h from K if it weakly orders the maxichoice contractions. Minimizing loss of version 2 damped informational value establishes that the mild contraction removing h from K is the weakest of the optimal options. Hence, if the Rule for Ties I favor that recommends choosing the weakest of the optimal options is adopted, the mild contraction is recommended for choice.

Rott complains that version 2 damped informational value is constructed in order to guarantee that in contraction, loss of informational value is minimized and ties are broken by the Rule for Ties. He complains about the lack of an independent motivation. Independent motivation is needed because Rott thinks that taking the minimization of loss of type 2 damped informational value as conforming to the principle of informational economy is misleading.

My proposal requires assessments of informational value to be weak orderings of the domain of potential contractions that satisfies weak positive monotonicity and extended weak positive monotonicity. These conditions are satisfied by all explications of assessments of informational value of which I know including the widely used measures of information based on probability measures. If Rott insists that a probability-based measure be used, then damped informational value of either type is not an assessment of informational value. A rose will smell just as sweet by any other name.

Whatever its name, damped informational value (whether version 1 or 2) weakly orders all the potential contractions in a manner that guarantees that suspending judgment between optimal contractions is also optimal. This constraint on assessments of the utility of potential contractions represents an important intellectual value. Assessing potential contractions in this way permits the use of a secondary criterion for breaking ties that recommends suspense in such cases. The rationale offered by Pagnucco and Rott fails to supply a weak ordering among all contractions to be optimized. It is not decision-theoretic. This leaves all weak orderings satisfying the damping requirement. There are weak orderings other than versions 1 and 2 that satisfy the damping requirement, but these two are the only two whose properties have been seriously explored (except by Horacio Arló Costa in his chapter for this volume). I think that version 2 damped informational value should be favored over version 1 because of the unattractive implications of version 1 for statistical applications as explained in *Mild Contraction*.

Rott rightly points out that I failed to provide a consistent definition of version 1 damped informational value in *The Fixation of Belief and Its Undoing*, *For the Sake of the Argument*, or *Mild Contraction*. Pace Rott, however, consistent definition can be given without serious modification of my discussion in *Mild Contraction* or any of the earlier references.

Two saturatable contractions removing h from K are *h-equivalent* if and only if they are intersections with K with different elements of the subset $S(\sim h)$ of U^*_K all of whose members entail $\sim h$ and the same subset of the subset $S(h)$ of U^*_K all of whose members entail h.

Every potential contraction K^* removing h from K is the intersection with K of a nonempty subset R of $S(\sim h)$ and a subset R^* of $S(h)$ that may or may not be empty. Hence every such contraction is the intersection of *h-equivalent* saturatable contractions removing h. Each saturatable contraction formed by intersecting K with a singleton from R and all the elements of R^* is h-equivalent with every other such saturatable contraction, and the intersection of all such saturatable contractions yields the contraction K^*.

The weak version of the intersection equality condition on page 125 of Levi 2004 should be amended to require that the set S of saturatable contractions be h-equivalent. Similarly, in the damping constraint on page 128, the set T of saturatable contractions removing h from K should be restricted to h-equivalent ones. Finally, the characterization of damping version 1 on page 130 should require that the set T be restricted to a set of h-equivalent saturatable contractions removing h from K.

These modifications take care of Rott's objection. I thank him for pointing out the error so that it could be corrected.

Where does this leave us? Imposing the damping requirement on weak orderings of *all* potential contractions in the dual ultimate partition does have an independent rationale. It allows suspense among optimal contractions to be optimal. This does not single out minimizing loss of version 2 damped informational value. But version 1 is safely eliminated for reasons stated in *Mild Contraction*. There are other alternatives. Pending closer examination of their doubtful prospects and once criticisms such as Hansson's are addressed, I think version 2 remains the utility function to optimize in contraction. And it rationalizes mild contractions.

CONTRACTION AND INFORMATIONAL VALUE: REPLY TO ARLÓ COSTA

I have little to say in response to Horacio Arló Costa's survey of our joint work and new frontiers that he is pushing except to acknowledge my gratitude for his collaboration. The following remarks take note of a few differences in emphasis in our attitudes that may be summed up by saying that Horacio is, among other things, an earnest builder of bridges where I tend to a perverse argumentativeness.

Horacio Arló Costa's chapter begins with a survey of work that appears in my *Mild Contraction* and our recently completed collaboration. *Mild Contraction* provides a decision-theoretic rationalization of the form of contraction Arló Costa calls "core contraction" and I called "impalpable contraction" together with a decision-theoretic rationalization of Arló Costa's "standard" contraction corresponding to my mild contraction. Arló Costa provides some consideration of the possibilities for forms of contraction other than impalpable contraction and mild contraction that satisfy the minimum requirements imposed on assessments of informational value. And he discusses axiomatizations of impalpable and mild contraction including his own demonstration of a complete axiomatization of mild contraction.

Mild contraction satisfies antitony. This has led S. O. Hansson and Hans Rott to register some doubts concerning mild contraction or severe withdrawal. Arló Costa adopts a neutral stance on the same issue. As I explained in my reply to Rott, I have little difficulty with antitony given that the range of potential contractions is restricted by a basic partition and minimal belief state.

I have been allergic to Grove diagrams because historically they have been an obstacle to understanding that assessments of losses of informational value must take into account valuation of all potential contractions from K that remove A and not merely maxichoice contractions that do so. Although Arló Costa clearly appreciates this point when developing his shells of informational value and cannot be charged with the oversight, I have acquired a certain distaste for geometric diagrams in representing belief changes.

Arló Costa has pushed on to new frontiers in discussing iteration of contraction in connection with mild contraction and under the assumption that assessments of informational value are held fixed. He uses his machinery (including, I must concede, his shells of informational value) to good effect in exploring the ideas of others and raising questions about the prospects for iterated mild contraction.

For my part, accounts of iterated contraction and iterated revision in the context of belief change are not very promising. The reason is that in belief change there is too much opportunity for relevant changes in the contextual parameters to take place so that simulation of the requirement of "categorical matching" becomes utterly inapplicable. Iteration of revisions and contractions also plays a role in the analysis of suppositional or conditional reasoning.

In that setting, it is plausible to suppose that the relevant contextual parameters remain fixed.

WHAT DECISION? REPLY TO KAPLAN

In cognitive decision problems characteristic of induction, the goals are purely epistemic and the options are decisions to *change* one's state of full belief or state of knowledge by expansion. The choice is deciding what to (come to) believe in the sense of a doxastic commitment. But the agent *X* cannot decide what to (come to) know. *X* has no control over the truth-value of what *X* decides to (come to) believe.

One cannot always control whether one fulfills doxastic commitments either. So the inquirer *X* does not decide what to believe if that means a decision to have certain doxastic dispositions such as a "willingness to bet all or everything on the truth of *P*" or to manifest such dispositions. Acquiring such dispositions and abilities is part of the process of fulfilling doxastic commitments and may require therapy and technological fixes.

There is another sense of "deciding to know that *P*." One might employ some criterion to decide whether a certain condition is satisfied. Does my current attitude toward *P* qualify as knowing that *P*? Perhaps by checking on the features of my current attitude, I can decide. But such deciding is not deciding between expansion strategies.

Mark Kaplan seems to be referring to my discussion of deciding what to believe in the sense of deciding how to change beliefs, but much of his commentary raises problems for a view that I do not recognize as my own concerning deciding what to know where it is not always clear whether "deciding" is in the first or the second sense.

This does not mean that Kaplan and I do not disagree and in substantial ways. Kaplan is a pedigree epistemologist as note 12 suggests and is a radical probabilist. Radical probabilists are committed to standards for serious possibility and, hence, to states of full belief or knowledge. But radical probabilists commit to a Parmenidean epistemology that insists that such standards are incorrigible and incorrigibly skeptical. I need not repeat my dissent from this view.

Kaplan's objection to my view, when freed of the misunderstandings to which I have just gestured, contains a serious point that is relevant to his support of radical probabilism.

I have for a long time maintained that the common feature of the proximate goal of decision making in all contexts calling for deciding among expansion strategies *ought* to be a balance between the aim of avoiding the importation of error and the aim of maximizing valuable information. The assessment of valuable information may require taking into account moral, political, economic, and other practical objectives as well as theoretical considerations, such as explanatory power and simplicity. However, this assessment is autonomously theoretical or epistemic in the sense that the assessment must be constrained by the purely epistemic constraint that if adding *h* to *K* results in a belief state that entails the belief state obtained by adding *g* to *K*, the informational value of the first expansion cannot be less than the informational value of the second expansion. (This is what I now call "weak positive monotonicity," but its provenance goes back at least to my "Information and Inference" of 1967 – see note 31 and the text on which it comments.)

Kaplan translates this epistemic autonomy thesis into an "account of how to decide what you know that will (1) respect the purely epistemic character of the considerations that determine whether a proposition deserves a place in your corpus of knowledge." Let me exercise charity here and rephrase this claim as a decision as to what you decide to come to believe and modify the final phrase by inserting "new" before "corpus of knowledge."

Thesis (2) when charitably reconstrued asserts that on deciding to add *h* to one's initial state of full belief, one undertakes the commitment to "bet the house" on the truth of *h* should the opportunity to take up such a bet arise. Such a commitment is entailed by my epistemological infallibility thesis that mandates commitment to incorporate *h* into one's standard of serious possibility or state of full belief

However, *X* may not be "willing" or "prepared" to bet the house in the sense of having dispositions and abilities that fulfill the commitment undertaken. This observation parallels the observation that even though *X* is committed after adding *h* to judging true all logical consequences of *K* and *h, X* will not have all the dispositions and abilities needed to fulfill this commitment. *X* will not be logically omniscient.

Moreover, *prior* to adding *h*, *X* is not even committed to bet the house.

Now given (1) and (2) as modified so as to represent something like the view I endorse, one might still worry that one cannot justify coming to full belief in the truth of an extensive number of extralogical propositions as my inductivism (thesis (3)) requires.

Prior to expansion, the falsity of *h* is a serious possibility. Expanding *K* through adding *h* is optimal (let us suppose for the sake of the argument) with respect to the proximate epistemic aim of obtaining new valuable and error-free information regarding a specific question. But agent *X* has many other value commitments: to maintain and support life and limb, to promote peace, and so on. Prior to expansion, *X* would not accept certain bets on the truth of *h*, although *X* might accept others. But prior to expansion *X* recognizes that after expansion *X* will come to full belief that *h* and will be prepared to bet the house. This is a *consequence* of the expansion.

If *X* were concerned solely with promoting the epistemic goal of obtaining the new, valuable, and error-free information, this consequence would be irrelevant because its value would be discounted in deliberation. What I think Kaplan's objection ought to have been is that *X* may refuse to ignore *X*'s other value commitments and focus exclusively on *X*'s epistemic goals. The epistemic goals by this argument will virtually always come into conflict with the other values of the agent. Either the conflict will go unresolved, in which case expansion will quite likely not be recommended, or it will be resolved in a manner that will in general avoid recommending anything other than remaining with *K*.

This objection is directed at views I have endorsed and needs to be taken seriously. Even so, I do not think the objection is telling.

Agents with diverse value commitments do not necessarily recognize them as being in conflict in a way that calls for inquiry aimed at resolving the conflict. To mandate conflict, the possibility of conflict must be recognized as serious and not merely logical.

Suppose that expanding by adding *h* to K is warranted relative to epistemic goals. *X* would not be committed beforehand to betting the house, *X*'s life, or anything else of paramount value for a paltry benefit.

X may also be convinced at the time that no such bet will be offered. In that event, *X* will not recognize *X*'s moral, practical, and so on value commitments to be in conflict with *X*'s epistemic interests. *X* may be justified in expanding. Perhaps afterward *X* will be confronted with a bet on *h* incurring the loss of house or life if *h* is false. This will warrant revising *X*'s initial view that no such bet would arise. But it will not, in general, warrant withdrawing *X*'s full belief that *h* and *X*'s new commitment to the view that no risk is being faced.

This account presupposes that we often rule out the possibility of conflict arising between goals and values we are currently pursuing and other goals. The important argument suggested by Kaplan concludes that we should not do so. But if we do not, we would suffer from a moral equivalent of attention deficit disorder.

Ruling out logical possibility of conflict as a serious possibility is, of course, subject to modification in later experience and inquiry. This is true whether or not the goals being pursued are epistemic or practical. We must always keep in mind that sometimes genuine conflicts arise between the pursuit of epistemic values and other values. I insisted on this even in *Gambling with Truth* (pp. 18–19).

My contention is that Kaplan's argument fails to preclude endorsing the autonomy of epistemic value, epistemological infallibilism, and inductivism. What it does point to is the importance of recognizing the kind of value pluralism that John Dewey emphasized and that I sought to promote in *Hard Choices*.

COMMITMENTS, IDEALS, AND IDEALIZATION: REPLY TO SVEN OVE HANSSON

Sven Ove Hansson reports correctly from my "Rationality and Commitment" that I distinguished between three states of a commitment to attitude: a state of full belief, a state of credal probability judgment, and a state of value commitment. In writing this, I was doubtlessly idealizing in the sense of simplifying. Insofar as such states may be varied while keeping the others fixed, the credal state needs to be replaced by the agent's confirmational commitment, which is representable as a function C from potential states of full belief to credal states and characterizes the inquirer's commitment to adopting the credal state B that is the value of the confirmational commitment when the inquirer's state of full belief is K. Value commitments too should be able to change without changes in confirmational commitments or states of full belief. I conceive of the three attitudinal components as independently variable. In introducing them, I was not concerned to draw the fact-value distinction.

Indeed, many features of the fact-value distinction are obscure to me. Thus, value judgments such as probability judgments are neither true nor false. Full beliefs carry truth-values. So on which part of the divide between judgments of fact and judgments of value do probability judgments lie? No matter how one wishes to settle this and related matters, confirmational commitments (but not credal states) may be changed independently of states of full belief.

In any case, these states are states of commitment on my view and not states of performance. This contrast is not between what the inquirer *actually* fully believes and what the agent is *committed* to fully believing, what the inquirer actually would regard as probable to varying degrees were the inquirer in such and such a state of full belief and what the inquirer is committed to judging the appropriate probabilities to be were the inquirer in that state of full belief or between the agent's actual valuations and the valuations the agent is committed to having. On my view, X's commitments are just as "actual" as X's performances. Indeed, X's performances could not be understood as performances actual or otherwise unless they are interpreted as successful or failed attempts to fulfill commitments. In the absence of such construal, X's utterances, doings, and dispositions thereto lack intentionality.

Hansson suggests that my view makes sense if we focus on ideal agents with transfinite cognitive capacity but not if we focus on agents with limited cognitive capacity of which they make rational use. I disagree.

Like everyone else, I simplify by supposing that in the situations to which my proposals are to be applied, certain factors may be ignored as irrelevant for the purpose at hand. I also concede that I consider ideals in the sense of perfect rationality. But I insist that flesh-and-blood rational agents *in real life actually* fully believe *in the sense of doxastic commitment* all logical consequences of their full beliefs. X does not, of course, believe the consequences of X's beliefs in the sense of doxastic performance.

If X is committed to believing that h and all its logical consequences and we discover that X fails to acknowledge the truth of some proposition g deducible from h, it is commonly held that *either* X should fully believe that g or X should give up belief that h. I insist that the latter alternative should be ruled out. If X is committed to fully believing that h, X is committed to fully believing that g. X needs a good reason to justify changing X's doxastic commitments abandoning X's doxastic commitment to full belief that h. Absent such a good reason, X *should* retain the commitment and *should* improve X's performance by assenting linguistically and behaviorally to the truth of g when the occasion requires and costs and abilities permit. X should change X's performance but not X's commitment.

There is no insistence that X should do something X cannot do. But X is quite capable of undertaking to improve X's capacities to fulfill X's commitments. A doxastic commitment resembles a vow to chastity that one makes knowing full well that one cannot meet its requirements perfectly. Making the vow does, however, incur the obligation to do the best one can and in addition to improve one's capacities, costs and opportunities permitting.

Finally, I deny the "ideal of rationality" recommending that a rational subject should assign a definite (determinate?) probability value to each statement about the world less than 1 unless the statement is a logical truth. I am not alone. Even Bayesians with impeccable credentials, such as De Finetti, do so. The opposing view with its commitment to a Parmenidean epistemology has become popular in the last third of the twentieth century thanks to the influence of of R. C. Jeffrey and his acolytes.

I also deny that attributing extralogical full belief to X is merely a simplification of X's attitude when X judges the proposition to be probable to a degree near to but distinct from 1. As I have repeatedly stated, X's state of full belief is *X*'s standard for serious possibility where the space of serious possibilities generates the algebra over which *X*'s credal probabilities are defined.

Hansson charges me with wanting to have it both ways when choosing between levels of idealization where the highest level is the Bayesian level as R. C. Jeffrey understood it and the next highest allows for full beliefs in extralogical propositions as tolerable approximations.

I do not consider myself as choosing between these levels of idealization at all. Neither idealization looks wholesome to me for any purpose, philosophical or practical.

The function of principles of rationality is to characterize how commitments to full belief, probability judgments, and value judgments are generated by *X*'s attitudinal performances (dispositions to behavior and their manifestations).

Hansson closes with a reference to a brief discussion of commitment and determinism in my "Commitment and Change of View." I am afraid that I failed to make clear the problem with which I was concerned. I was not trying to solve the metaphysical problem of free will. I support some form of soft determinism that allows that agents can be in control of some of their behavior even though that behavior is causally explicable under some description or other by antecedent factors. I even agree that sometimes we can change our dispositions to behavior "at will." But the scope for control of belief would be seriously circumscribed according to such a view, just as Hume indeed took it to be. I am concerned that the inquirer *X* have control over *change* in belief in the sense of change in doxastic commitment. In making such a change, *X* is also adding or removing something to or from *X*'s initial belief state. Such a change may be accompanied by a public declaration or some action or it might be accompanied by some soliloquy. Such activities can be under *X*'s control according to soft determinism, whether they are described as performances or as undertakings of doxastic commitments.

NATURALISM AND THE MIND: REPLY TO HINZEN

In his excellent, challenging, and passionately argued chapter, Wolfram Hinzen argues that the acquisition of grammatical competence is the acquisition of a "system of knowledge" without any effort at problem-solving inquiry resulting in propositional knowledge but rather by a process of maturation that depends only to a "minor extent" on environmental feedback or cultural difference. The only way I can understand the claim that grammatical competence is a system of knowledge is as a misleading way of stating that grammatical competence is a skill or ability.

The claim is misleading in two respects: First, even though a skill may be described as "knowing how" to do something, it is not a system of knowledge in the sense of a theory. The system involved is rather a system in the sense in which the auditory system is. Such a system is an organic precondition for the ability to exercise the skill of listening to music, speech, and other sounds. And the grammatical system may be a similar precondition for the ability to express commitments to full belief, probability judgment, value judgment, and conclusions of deliberation. My grammatical and linguistic experience suggests to me that calling such systems systems of knowledge is a misleading neologism.

Second, calling grammatical competence "knowledge" suggests that the skills are intentionally characterized. Perhaps it is unfair to suggest that Hinzen misleads here since he is quite open about insisting that intentionality is simply a fact of nature. But I see no more grounds for thinking of grammatical competence as exhibiting intentionality than for thinking that the auditory system or the respiratory system does. And concluding that investigations into early infant discrimination between human and inanimate objects establish the presence of intentionality in early infants without reference to how well the infants are doing in fulfilling attitudinal commitments in accordance with standards of rational health is as convincing as my judging that my dog exhibits shame, defiance, and affection on the basis of his behaviors.

There is no doubt that we are wont to attribute attitudes to animals and infants even when it is difficult to see such attitudes as commitments. The same is true of subconscious motives and attitudes. Akeel Bilgrami has suggested in the case of the subconscious a modification of my proposal to characterize intentionality of behavior as failed or successful attempts to fulfill commitments. Subconscious motivations are taken to be impediments to fulfilling commitments. As such they need not be intentional at all. The contentful descriptions of them occur due in large measure to what Bilgrami calls our "epistemic weakness." We lack an explanatorily adequate characterization of these systems and, instead, refer to them in ways that indicate the commitments they are frustrating. But insofar as our aim is naturalistic explanation, such descriptions should be replaced by neurophysiological or other physical characterizations that contain no whiff of intentionality or normativity.

The study of early childhood cognitive development, it seems to me, is of importance not as part of a theoretical or explanatory enterprise but as part of an effort to improve clinical skills in treating various learning disorders. My claim that it is not part of an explanatory study is predicated on the assumption that its theoretical apparatus reveals the same signs of epistemic weakness that Bilgrami notes in discussions of subconscious motivations in psychoanalysis. If such investigations are to become of theoretical as opposed to clinical importance, they would have to look to the physical and neurophysiological processes to provide the explanatory resources that should replace the appeals to dispositions and abilities (also known as "competences") and to intentionality. There is nothing wrong with beginning with placeholders for theoretically adequate notions in scientific inquiry as long as one acknowledges the epistemic weakness and the importance of overcoming it.

My worry about Hinzen's case for the natural mind that does not change once it has matured is that it is the product of a view that fails to recognize the need for such acknowledgment. Such a view makes a mystery out of dispositions and abilities, converting them into occult powers. That is too great a price to pay in order to declare oneself a naturalist.

COMMITMENTS AND RATIONAL HEALTH: REPLY TO BILGRAMI

Akeel Bilgrami and I have carried on a longstanding conversation that has been an abiding benefit to me. Our disagreements are Trotskyite when compared with our differences with popular positions advocated by those interested in psychology, rationality, and the relation of psychology to the natural sciences. But for me such differences are fascinating. Bilgrami's first-rate discussion of the subconscious in psychoanalysis and its relation to our shared emphasis on the centrality of normativity to the intentionality of the psychological and to agency brings out some of these differences very nicely.

Two kinds of change in view are relevant to the study of inquiry: change in commitment and change in performance so as to fulfill a commitment.

In the case of full belief, a change in commitment is a change in the state of full belief or doxastic commitment $\boldsymbol{K}$. Agent X may undertake the change explicitly by some linguistic or other public declaration, or the undertaking may not be expressed publicly at all. In any case, in that state X is committed to believing the logical consequences of the beliefs to which X is committed according to that state, to believe that X believes that h if X believes that h and to believe that X does not believe that h if X does not believe that h.

X lacks logical omniscience and capacity for self-reflection that would be required to fulfill all the commitments thus generated. Nonetheless, X is obliged to fulfill the commitments when X is able and costs permit and to support efforts to enhance X's capabilities and reduce costs when this is feasible. Such changes in performance are linguistic and behavioral acts or acts of reflection that carry intentionality or they are changes in dispositions to such performance. I have suggested that when they are attempted fulfillments of doxastic commitments, they are understood as bearing contents that are relevant to ascertaining how successful *X* is in doing so.

Bilgrami correctly points out that subconscious dispositions cannot plausibly be understood as attempts to fulfill doxastic or other attitudinal commitments even though they too are often characterized in intentional terms. He suggests interestingly that they are often described intentionally but only as a reflection of our epistemic weakness. Impediments to fulfilling meaningful commitments are recognized, but they cannot as yet be integrated in an adequate biological and physical scheme. They are represented as carrying content as a means for marking commitments whose fulfillment they frustrate. When such impediments are recognized, a decision has to be made as to whether to modify commitments so as to accommodate the presence of the disposition as a partial fulfillment of the modified commitments or whether to engage in efforts to remove the disposition. Bilgrami notes that uncovering the dispositions in the first place and, when it is judged appropriate, removing them may call for a "technological" fix. Bilgrami's characterization of subconscious dispositions is a useful extension of the placeholder approach to dispositionality proposed by Morgenbesser and me a third of a century ago.

Bilgrami and I agree that the principles of rational full belief (rational probability judgment, preference, etc.) characterize the commitments generated by an undertaking. An agent *X* ought rationally to believe fully that *X* fully believes that *h* if *X* fully believes that *h*. For that reason, if *X* is committed to fully believing that *h*, *X* incurs the commitment to fully believe that one fully believes that *h*.

We also agree in understanding the intentionality of propositional attitudes in terms of commitments. In addition, the status of subjects as agents that are evaluated with respect to whether they fulfill or fail to fulfill doxastic and other attitudinal commitments is a matter for normative reflection and adjudication. And we agree in denying that psychology – insofar as it considers agents bearing such attitudes – can yield successful and satisfactory scientific explanations, although it can be a clinical activity or, as Bilgrami suggests in the case of psychoanalysis, a technology.

Even so, we have some differences. On my view, *X* does what *X* does intentionally and in so doing incurs an obligation to fulfill the commitment that is undertaken in thought and deed. But the obligation is not a moral obligation such as *X* would incur by promising *Y* that *X* will do something. *X* need not be prepared to accept criticism from *Y* if *X* fails to fulfill *X*'s doxastic obligations.

Undertaking a doxastic commitment is more like making a religious vow than making a promise. The religious declare that the vow is to God. I think that the vow is made religiously without being made to anyone. *X* is responsible to no one – not even to *X* – for fulfilling the commitment. If *X* makes a promise *X* is convinced that *X* cannot fulfill, *X* has promised fraudulently. Not so with religious vows or doxastic commitments. Just as religious vows are vows that *X* knows that *X* cannot fulfill perfectly and yet is not fraudulent for making the vow, so too commitment to believing the logical consequences of one's beliefs is no fraud even if *X* is perfectly aware that the undertaking cannot be perfectly fulfilled. *X* should fulfill the vow or commitment to the best of *X*'s ability and for the rest seek to improve *X*'s capacity – costs, opportunities, and abilities permitting.

If *Y* claims that *X* has failed to believe fully a logical consequence of full beliefs to which *X* is committed, *X* is not obligated by *X*'s commitments to accept the criticism from *Y*. If *Y* is mistaken, *X* has no such obligation. And if *Y* is correct, *X* is in a state of rational incoherence. How can *X* be prepared to accept *Y*'s criticism for *X*'s failure then? *X* can be committed to such readiness but obviously may not be in a position to fulfill the commitment.

According to Bilgrami, the normativity of doxastic commitments is grounded in "reactive attitudes." Praise and blame, reward and punishment, and holding *X* responsible is appropriate. This makes sense for the normativity of moral commitments. Failure to fulfill them can justify such reactions. But failure to fulfill the requirement that one should believe the logical consequences of what one believes or that one should believe that one believes given that one believes it does *not* justify these responses. If *X* fails to fulfill *X*'s doxastic commitments, *X* has failed to be rationally coherent. This I suggest is rather like being in a poor state of rational health. To suggest that *X* is in such a poor state ought not to call forth the reactive attitudes unless it can be shown that holding *X* accountable for failure to fulfill the commitments generated by full beliefs in accordance with principles for rational full belief will effect a cure. Bilgrami and I agree that the intentionality of the attitudes is bound up with normativity. But the normativity derives, on my view, from the commitment to achieving a *mens sana* and not from the reactive attitudes. I see these obligations as akin to striving to achieve an unachievable ideal characteristic of certain types of religious undertakings or attempts to attain perfect health.

Placing the reactive attitudes center-stage puts too much emphasis on the more misguided aspects of Kantian morality. According to Kant, moral commitments, are, indeed, rational commitments, and the obligations generated are in both cases dictates of reason. The propriety of reactive attitudes is not to be assessed in terms of the consequences of holding agents responsible for successful compliance but is seen, as Bilgrami sees it, as constitutive of the normative idea of agency.

On this view, failures of morality are easily confused with failures of rationality. And this leads to the idea that rational agents are not committed to believing the deductive consequences of what they believe. To claim that failures of logical omniscience are moral failures has the air of absurdity about it.

By the same token, considering deviations from some moral norm to be a mark of rational incoherence rather than immorality renders opening up one's mind to the point of view of the alleged deviant by modifying one's moral commitments more difficult than it would be if the alleged immoralist is taken to be rationally coherent. To regard *Y* to be an advocate of immorality may be taken to be a sign of moral sickness or to be an honest difference of a opinion. To regard *Y*'s advocacy of immorality to be rationally incoherent takes the second option off the table.

On the view I favor, standards of rationality are, on the one hand, quite minimal. They are weak in the way logic is supposed to be. On the other hand, satisfying these standards is very demanding and, indeed, is well beyond human capacity. In my judgment, the understanding of doxastic and other attitudinal commitments as akin to the religious pursuit of an ideal of rational health characterizes the normativity of attitudinal commitments quite well.

Of course, there is an important respect in which *X* ought to be prepared to accept criticism for failure to fulfill *X*'s doxastic and other attitudinal commitments. Self-criticism is essential to deliberation and inquiry aimed at promoting the subject's goals. But self-criticism and openness to advice from others conceived as integral to deliberation is not characteristic of reactive attitudes of praise and blame and accountability.

If Oblomov lacks agency, he does so not because he lacks a first-personal as opposed to a third-personal point of view but because he lacks a point of view. There are, perhaps, useful distinctions to be made between *X*'s attitudes concerning *X*'s attitudes and *X*'s attitudes concerning *Y*'s attitudes or between *X*'s view of *Y* as a system in nature and as an agent. But these are distinctions between aspects of *X*'s point of view.

To be sure, *X* has the commitment to fully believe that *X* believes that *h* if *X* is committed to believe that *h*, as positive introspection requires, but has no commitment to believe that *Y* believes *h* when *Y* is committed to believe that *h*. This is not a first-person privileged access. There is no such thing. It is a first-person commitment. Bilgrami and I agree about this. In the past I have sometimes misleadingly called the first-personal commitment a first-personal point of view. It is first-personal in the sense that it is *X*'s point of view about

X's current attitudes and contrasts with *X*'s point of view about *Y*'s current attitudes and the attitudes of *X* and *Y* at other times. I do not discern in this a difference between *X*'s first- and third-personal points of view about *X*'s current attitudes as seems implied when it is alleged that though Oblomov has a third-personal point of view about himself, he lacks a first-personal point of view.

I do not agree with Bilgrami that an agent cannot see him- or herself "as a product of causes and the trajectory of predictions." In a context where *X* is deliberating as to what to do, *X* cannot coherently have information about the causal influences on him that warrant *X*'s predicting what *X* will do. Deliberation does indeed crowd out prediction. But that does not mean that *X* cannot be convinced that his or her decision will be the product of universal causation.

Bilgrami's "psychiatry driven ideologues" who claim that human actions are the product of *coercive* causes may coherently do so each from his or her ideologically driven point of view. The ideologues are wrong-headed but not incoherent. Bilgrami's reaction to such views is perfectly appropriate. The decision as to whether agents are criticizable for their actions depends on the value commitments of those who make the decision. Bilgrami's contention that the question as to who or what is an agent has an ineradicably normative or evaluative component is correct. But the correctness of his claim has nothing to do with a nonexistent distinction between first- and third-person perspectives or with the centrality of the reactive attitudes.

HOW ADAPTIVE SHOULD OUR RATIONAL PREFERENCES BE? REPLY TO RABINOWICZ

As usual, Wlodek Rabinowicz has written a carefully and clearly argued brief for a point of view for which I have little sympathy.

I have argued that persistent bookies do not assure the cogency of diachronic pragmatic arguments because they do not show that violators of acyclicity or reflection are going to choose options that are dominated by others available to them. Rabinowicz responds by modifying the predicaments on which the diachronic arguments are based.

Suppose the agent *X* with cyclic preference over *x*, *y*, and *z* is confronted with a choice among these three prizes. There is no optimal option in the three-way choice. *X* should eliminate the cyclicity from *X*'s preference so that *X*'s preference can function as a guide in optimizer *X*'s deliberations. Cyclic preferences are irrational precisely because *X* cannot choose rationally in some decision problems. Were *X* confronted with a three-way choice among *x*, *y*, and

z, X could not follow the policy of choosing an option that is V-admissible in the sense that it is optimal according to some permissible ranking and, indeed, could not follow the slightly different policy of choosing an option that is maximal in the sense that no option is strictly preferred to it. I am convinced by this argument that cycles should be avoided. Rabinowicz's argument seems far less compelling.

Consider Rabinowicz's sequential choice problem as illustrated in Figure 19.2. X faces a single synchronic and unified decision problem at the initial node. According to the situation, X begins with an initial endowment of wealth x. As X is offered a "trade" where X receives $Y - \varepsilon$ for x. These are X's two options. If X accepts the trade, X is certain that X will face future trades that X will accept, leading to X's ending up with $x - 3\varepsilon$. If X refuses the trade, X will end up with $z - 2\varepsilon$. In spite of X's cyclic preferences, X will choose to accept the trade.

In this case, X chooses the best option available to him or her. There is no irrationality in this as far as the argument goes.

Rabinowicz proposes therefore to consider a scenario where X also has an additional option: "to decide on the whole temporal sequence of his or her actions." However, X does not deliberate on the sequence as a whole.

A possible way to understand Rabinowicz's suggestion is that X has control *at the initial node* over which of the eight paths X will choose. But X deliberates in a "disunified" way so that at each node he or she deliberates between the sell/don't sell options available then.

If X refuses to consider all the options that are available to X according to X's beliefs and goals, X's deliberation is irrational. Indeed, this is so whether or not the options that are not considered dominate the one chosen from the options that are. Such disunity is to be avoided. This is so whether X is offered a set of gambles at the same time or is offered a sequence of options where X regards X to be in control of the path X will take. So let X deliberate between the eight paths.

These paths have four distinct payoffs that produce a preference cycle. And the preference cycle precludes there being maximal or V-admissible options among the eight available options. Each of the available options dominates another available option and is dominated by yet another. But the argument to this result is not diachronic. X's acyclic preferences are shown to be incoherent by a *synchronic* pragmatic argument.

I have not been quite faithful to Rabinowicz's depiction of the scenario. Rabinowicz invites us to consider a case where X has as an *option* "to decide on the whole temporal sequence of his or her actions." So X is to *choose* between a unified approach and a disunified approach. If this means anything, it means that X can choose whether to control X's future choices or to renounce such control.

Figure 19.2 should then be modified. A preinitial choice node should be added. That node contains two options. One is to the eight-option predicament and one to X's Figure 19.2 decision problem. At the preinitial node, X is certain that if the Figure 19.2 decision problem is chosen, X will end up with $x - 3\varepsilon$. If X chooses the other way, X faces incoherence. Assuming that such incoherence should be avoided, X should end up in the Figure 19.2 predicament. Once more X does not choose a dominated option.

In spite of this negative verdict on Rabinowicz's response to my comments on Skyrms's and his diachronic incoherence arguments, a legitimate point lurks behind his discussion.

My synchronic argument against cyclic preferences argues that cyclic preferences prevent rational choice whether maximality or V-admissibility is the standard for rational choice. Because of this unfortunate consequence, preferences ought to be modified so as to conform with acyclicity.

X's having cyclic preferences in Rabinowicz's Figure 19.2 predicament leads to X facing a loss no matter which of the options available to X is chosen. If X wishes to "keep on the sunny side," X might "adapt" X's preferences.

My argument and Rabinowicz's can then be placed on an equal footing by noting that both invoke arguments for adapting preferences. The problem raised has nothing to do with diachronic versus synchronic rationality. Rather, it concerns the kinds of bad consequences whose avoidance requires principles of rationality rather than principles of morals or prudence.

So reconstrued, Rabinowicz's discussion has raised an interesting point.

For starters, I concede that none of the so-called pragmatic arguments for principles of rationality are purely pragmatic. Dutch book arguments for the requirement that betting rates conform to the calculus of finitely additive probability presuppose state-independent preferences and that preferences among gambles constitute a weak ordering. There is no free lunch.

When preferences go cyclic, decision problems can arise where the agent cannot identify a maximal or V-admissible option. The most rudimentary requirements for deliberation are frustrated.

Few people enjoy being in a situation where they are going to lose, no matter what they choose. But such tragedies do occur without precluding rational deliberation. Changing preferences in such cases is not rationally mandatory. The decision maker retains rationality whether or not the change is made.

Frustrating minimal requirements for practical deliberation when such deliberation is called for is to be avoided at all costs. So, it seems to me, is requiring rational agents never to change their values and never to regret their prior probability judgments. These and other rigidities are the byproduct of diachronic pragmatic arguments. When it comes to logic, we should follow Aristotle rather than Hegel. Doing so prepares us all the better to address the Heraclitian and Hegelian flux.

PLACEHOLDERS AND STOPGAP EXPLANATIONS: REPLY TO PERSSON

A "surefire" disposition predicate "$D_{R/S}x$" appears in a universally generalized statement $(x)(t)[D_{R/S}x, t \wedge Sx, t \supset Rx, t]$ that may be paraphrased as claiming that if x is disposed at t to respond in manner R on trial of kind S and x is subject to a trial of kind S at t, x responds in manner R at t. The number of places in the disposition, trial, and response predicates may be more than indicated and the time at which the disposition is present may be related to the trial and response times in more sophisticated ways than indicated here. And surefire disposition predicates may be "multitrack" in the sense that they are characterized by programs that specify responses for different trials or inputs.

In his clear and interesting chapter, Johannes Persson points out rightly that on the view proposed by Morgenbesser and myself and subsequently elaborated by me in several ways, the universal generalizations in which problem-raising disposition predicates appear are stopgaps for candidate laws to serve in covering law explanations. The disposition predicates are "surefire" because the principles that link the presence of disposition with input and output are these stopgap universal generalizations. The disposition predicates are not themselves universal generalizations but placeholders for theoretical terms appearing in the promised covering laws. And the stopgap laws are not considered adequate for explanatory purposes as they stand. I am somewhat puzzled, therefore, by Persson's locating the difference between Elster's view and my own as between Elster's insisting that disposition predicates are not surefire and that lawlike generalizations are scarce while mechanisms abound.

As Persson says, if the disposition to have an adaptive preference is triggered, the disposition to have a counteradaptive preference is not – and vice versa. But to speak of a disposition to have a response of some kind does not make sense unless one specifies the kind of trial that triggers the disposition. Are the triggers of the two dispositions of the same type or are they different? If different, the dispositions are surefire. If the same, the "mechanisms are not dispositions but abilities."

A coin is a system where tossing sometimes leads to landing heads and sometimes to landing tails. The coin does not have a sure-fire disposition to land heads up on a toss. That is to say, it has the ability to fail to land heads up on a toss. It also has the ability to fail to land tails up on a toss. However, the coin does have the surefire disposition to land heads up or tails up on being tossed and the surefire disposition not to do both. We may then say that the possible outcomes of tossing the coin are landing heads up and tails up. When Elster speaks of the way things (might?) happen on a certain kind of trial, I conjecture that he is thinking of abilities.

In *Gambling with Truth*, I already suggested that attributions of chance might be construed as placeholders. I elaborated and corrected my proposals in *Enterprise of Knowledge* and contended that attributions of ability (as duals of disposition predicates) and of sample spaces of possible outcomes on kinds of trials (as conjunctions of dispositions and abilities) are to be approached in the same way. All of these placeholders are introduced to provide schemata for covering laws for use in stopgap covering law explanation showing why certain types of events are to be expected and how events are possible. Attributions of chances serve as placeholders for various types of statistical explanation.

One can imagine even more complicated combinations of the placeholding modalities. As I have pointed out elsewhere, not all types of attributions contemplated by Persson and others are acceptable as placeholders. Finkish dispositions are unacceptable qua problem-raising placeholders. The motive for introducing such placeholders is lacking. Still, if placeholding disposition predicates become integrated into theories and, in effect, qualify as explanatorily adequate theoretical terms, they may in the course of such inquiry come to exhibit the properties of finkishness. I cannot elaborate here.

As far as I can make out, my disagreement with Elster does not concern the abundance or paucity of covering laws. Elster seems to think that in the social sciences we often and, perhaps, typically cannot provide explanatorily acceptable covering law explanations. It is easy enough to supply placeholders for theoretical predicates in schema for covering laws. But in the social sciences, Elster seems to be saying – and I am inclined to agree – that there is not much hope of converting such placeholders into theoretical terms in explanatorily and empirically adequate theories. The difficulty is not that explanations of human behavior will not be forthcoming, but it is likely that, as inquiry proceeds, the categories used for describing and explaining such behavior will be so transformed that the original dispositions, abilities, or mechanism will not be digestible in the new theoretical framework.

Elster suggests that we resign ourself to the facts and accept what I have called stopgap explanations as the best we can have. I resist relaxation in order to declare the achievements of social science to be progress in theoretical explanatory science

Such relaxation endorses the use of mystery-raising dispositions. It licenses the acceptability of the kinds of explanations proffered by those who explain linguistic competence or performance in terms of the presence of faculties. It proclaims victory on behalf of naturalism by changing the standards for adequate explanation rather than by empirical discovery.

I resist the relaxation. Instead, I suggest leaving the explanation of human behavior to the biological and physical sciences and taking empirical work in the social and psychological sciences in a clinical direction focused on improving the functioning of human beings and their institutions. On this approach, empirical research in the social sciences will be directed by conceptions of rational, moral, and social health.

Index